“I OUGHT TO BE THY ADAM”: ASIMOV’S *I, ROBOT*, CREATION, AND PERSONHOOD

by

KATIE GOOGE

(Under the Direction of Carolyn Medine)

ABSTRACT

“I Ought to be Thy Adam”: Asimov’s I, Robot*, Creation, and Personhood* will examine the religious importance of artificial intelligence narratives by focusing on Isaac Asimov’s *I, Robot*. It will begin by examining the history of artificial intelligence, both in science and in fiction, and argue for the importance of Asimov’s work in these fields. It will focus on the creator/creation relationship in the *I, Robot* stories and examine the ways in which robots can be important participants in the debate over personhood and identity. This work will examine the ways in which Asimov creates a moral framework and how this ethical system interacts with questions of personhood and identity. It will argue that science fiction and *I, Robot* make important contributions to the religious discussion of creation and selfhood and that they are uniquely situated to play a role in future conceptions of artificial intelligence and identity.

INDEX WORDS: Isaac Asimov, I, Robot, Frankenstein, Mary Wollstonecraft Shelley, robots, artificial intelligence, personhood, ethics, science fiction, religion

“I OUGHT TO BE THY ADAM”: ASIMOV’S *I, ROBOT*, CREATION, AND PERSONHOOD

by

KATIE GOOGE

B.A., University of Georgia, 2017

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2017

“I OUGHT TO BE THY ADAM”: ASIMOV’S *I, ROBOT*, CREATION, AND PERSONHOOD

by

KATIE GOOGE

Major Professor:	Carolyn Medine
Committee:	Sandy Martin
	Christopher Pizzino

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

DEDICATION

To my mother, for patiently listening to my rants about robots. And to my father: I blame you.

TABLE OF CONTENTS

Page

CHAPTER

1 Pygmalion to Campbell.....1

2 The Question of Personhood.....27

3 Moral Subjects and Objects 65

REFERENCES 104

CHAPTER 1

Pygmalion to Campbell

Isaac Asimov's Robot stories, written between 1940 and 1950, are some of the most influential stories in the literary history of artificial intelligence.[1] They were written at an important moment in the development of AI and play a significant role in later popular and technological understandings of the concept. In the 1940s computer science was becoming a discrete field, the general culture was growing more and more fascinated with AI, and science fiction was a popular medium with a large audience. As Asimov's stories were being published, the vocabulary, philosophy, and scientific study of artificial intelligence were still forming. These original nine stories introduced new questions about personhood, creation, and ethics through the lens of these robots, and influenced the science fiction discourse surrounding these topics for decades.

The first step to understanding Asimov and his implications for religious studies is to understand the background of his work as well as its later impact. Asimov built on a long tradition of technological and literary conceptions of artificial beings. He took ideas born in such works as Mary Shelley's *Frankenstein* and Karel Čapek's *R.U.R.* and changed the conversation about artificial beings by asking new questions born out of his own experiences in the mid-twentieth century. These questions make it possible to move into a discussion of the stories themselves and the kinds of beings that populate them. It is almost impossible to determine the

[1] For the purpose of clarity, the term "artificial intelligence" or "AI" will also be used to refer to beings that can be retrospectively considered "artificial intelligences" under the contemporary definition, even if the phrase did not exist at the time in which the being referred to was written or created. This is due to the utility and intelligibility of the word, and the lack of real distinction between literary depictions of artificial intelligences before and after the term was coined.

exact personhood of the robots, and Asimov himself was unclear on the matter. However, by examining the different ways that the robots interact with other beings and their claims to personhood, as well as the way that Susan Calvin, the primary human character who interacts with robots and other people, it is possible to begin to understand the new paradigms Asimov was forming for relationship with a created other. Finally, Asimov's robots can be understood through an ethical framework. Under the Laws that he created, Asimov insisted on correct action by his robots, and the question of moral agency and the moral responsibilities of these robots and of humans' responsibility toward their creation is a live one throughout this text.

Asimov began his career at a time when technology was rapidly advancing and people's perceptions of the possibilities of technology were changing. The Second World War began and ended during the time period that these stories were written and published. Despite the technological horrors of World War II, Asimov continued to defend and believe in the promise of technology as a way to help humanity and even prevent conflict. He took his culture's questions and doubts about technology and change and attempted to combat them with a strict ethical model that he believed to be applicable to both creations and creators alike.

Asimov was a self-proclaimed skeptic, secularist, and rationalist, but his works represent a major shift in the way culture as a whole approaches the religious questions of creation, personhood, and the ethical treatment of the other. Asimov considered what gods the robots would believe in, how they would approach their creators, how they would be limited, and how humans should face their technological children. In the study of religion, ethics is frequently considered in terms of ideals that humans will almost never achieve. Asimov, however, combined a popular perception of correct ethics with the scientist's need for a code that can be simply explained and easily enforced, and so created the Three Laws. For all the ontological

questions that the robots raise, Asimov did not intend to create a metaphysics; he created a practical, behavior-based ethics for both humans and robots.

Before beginning to analyze the text itself, it is important to situate Asimov's stories in their scientific and literary context and show how this central position in the early conversation about AI has had a lasting impact on the field. Asimov's initial premise of an artificial being is by no means original. The idea of a human creation with a life of its own has a long history in myth, art, and literature, from the golem to Frankenstein's monster. Both organic and mechanical "automata" have always fascinated humans. Noreen Herzfeld, in her theological examination of artificial intelligence, argues that this impulse stems from the "larger question of what it means to be human[, which] is graphically posed by AI. The goal of AI is to create an 'other' in our own image. The image will necessarily be partial."[2] Indeed, despite centuries of effort in this direction, there have been no successful efforts to create a strong artificial intelligence. [3]

The first real attempts at constructing these beings began at the end of the Middle Ages, when people began to build automata that would imitate human behavior, often with the end goal of creating artificial life.[4] This led to an obsession with clockwork beings in the seventeenth and eighteenth centuries. Entertainers and scientists alike tried to create machines out of gears and springs that functioned like living organisms. Patricia Warrick describes "[a] menagerie of clockwork animals---tigers, horses, dogs, cocks, songbirds---[which] entertained and amazed the aristocracy with their ingenuity in simulating life and motion."[5] However as science continued to

---

[2] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), ix.

[3] Artificial intelligence in computer science is divided into the categories of "strong AI" and "weak AI." Strong AI currently exists only in fiction, and is an attempt to create an artificial being that is a full person, with consciousness and human-level intelligence. Weak AI is what most of the field focuses on, and it works to create more limited AI that can perform some tasks, but is not conscious. Self-driving cars and systems such as Siri are both considered weak AI.

[4] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 163.

[5] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 33.

develop, and it became clear that these clockwork automata were not and never would be full persons, they became mere party tricks rather than real scientific experiments.[6]

This failure of science, however, did not prevent the development of a rich literary and mythological tradition of artificial beings. From the Greek myths of Hephaestus' metal helpers and Pygmalion's statue, people have always been fascinated by stories of humans creating other beings. However, in these early stories, including the story of the golem and up to *Frankenstein*, this human creation of an artificial being was very different from the modern perception of robots and artificial intelligence. It was most often a spiritualized and unique event, where the product would be a single artificial being, often created with divine influence.

This view began to shift, however, with Mary Wollstonecraft Shelley's *Frankenstein*, which is widely considered to be one of the most influential stories of artificial creation ever written. In an 1816 ghost story competition with Lord Byron and her future husband Percy Shelley, she was inspired by the contemporary fascination with electricity and galvanism to write what many view as the first science fiction novel.[7] *Frankenstein* tells the story of a young and reckless scientist who discovers the secret to life and uses it to animate a body constructed out of parts of corpses. Upon seeing his creation brought to life, Frankenstein is repulsed and abandons his "monstrous birth." The monster, however, far from his shambling and grunting screen counterpart, soon learns to read and speak, forming complex analyses of John Milton's *Paradise Lost*, and drawing parallels between the text and his own existence. However, having been spurned by Frankenstein, the monster turns against the scientist and begins a vendetta ultimately resulting in the destruction of creator and creation alike.[8] Shelley's work is especially notable in

---

[6] Niran B. Abbas, *Thinking Machines: Discourses of Artificial Intelligence* (Hamburg: Lit, 2006), 34.
[7] Lucy Morrison and Staci L. Stone, *A Mary Shelley Encyclopedia* (Westport, Connecticut: Greenwood Press, 2003), 157.
[8] Mary Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012).

the history of thought on artificial intelligence because of two main features: its place as the first narrative of artificial creation by scientific means, and its use of religious imagery and themes to discuss that act of artificial creation, without giving into a religious or spiritual explanation.

Though earlier authors incorporated artificial life in their works, the combination of contemporary scientific theories with extrapolative narrative made Shelley's novel the first of its kind.[9] Though the scientific advancements Shelley was responding to did not pave the way for artificial life, her literary contribution continues to shape the popular perception of created beings and influence science fiction. During her life, England entered into the Enlightenment and Industrial Revolution, which were both characterized by a dramatic increase in technology and science.[10] New scientific discoveries changed the way people saw the world and led people to question old models of life and society. Rapid social change encouraged people to think radically new thoughts about the way the world could be. For the first time in many centuries, the lives of individuals were changing drastically and diverging from the modes of living established by their parents and grandparents. With these new technologies and new ideas, people also began to consider role of technology and what its limits should be.[11] While this question was pressing for many parts of society, Shelley's close association with the Romantic Movement likely influenced her treatment of this subject. The Romantics, including her husband Percy Shelley and their mutual friend Lord Byron, were largely atheistic, but still heavily emphasized the natural over the technological.[12]

---

[9] Isaac Asimov, *Asimov on Science Fiction* (Garden City, N.Y. : Doubleday, 1981) 183.
[10] Jasia Reichardt. "Artificial Life and the Myth of Frankenstein" in *Frankenstein Creation and Monstrosity*, ed. Stephen Bann, 135-157 (London: Reaktion Books, 1994), 137.
[11] Justo Gonzalez, *A History of Christian Thought, Volume 3*, 2nd ed (Nashville, Tennessee: Abingdon Press,1987), 335.
[12] Dewey, Joseph. "Romanticism." *Salem Press Encyclopedia* (January 2015): *Research Starters*, EBSCO*host* (accessed April 18, 2015).

This contemporary conversation about the role of science and technology is especially prominent in *Frankenstein* thanks to the frame narrative. The novel is presented as a confession from Victor Frankenstein to Walton, an exolorer who is undertaking an ambitious and dangerous journey through the Arctic. Walton's quest is often read as a parallel to the dangerous scientific endeavors that push the limit of the ethical and the possible, including, within the story, Frankenstein's own creation of the monster.[13] However, Shelley leaves this encounter ambiguous. Frankenstein counsels Walton to "[s]eek happiness in tranquility and avoid ambition, even if it be only the apparently innocent one of distinguishing yourself in science and discoveries. Yet why do I say this? I myself have been blasted in these hopes, yet another may succeed."[14] Though Frankenstein's own behavior is largely condemned by the narrative, Shelley stops short of discouraging all scientific progress, and the reasons for her condemnation of Frankenstein are complicated in ways reflective of her time period.

Part of this complexity is due to the religious themes and concepts that Shelley includes in her work. Asimov claims that "it is impossible to write science fiction and *really* ignore religion."[15] *Frankenstein*, at least, openly addresses one of the most important themes in Christian thought: the relationship between God the Creator and humanity as God's creation Shelley primarily engages with this concept through a dialogue with John Milton's *Paradise Lost*. This 1667 biblical epic, which relates Satan's rebellion in Heaven and subsequent fall to Hell, as well as the events of the first chapters of Genesis up to Adam and Eve's expulsion from the Garden of Eden.[16] Shelley likely chose this particular text because of her close connection to

---

[13] Brian W. Aldiss, *Billion Year Spree*, (New York: Doubleday, 1973), 22.
[14] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 157.
[15] Isaac Asimov, *Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995), 231.
[16] Henthorne, Susan. "Paradise Lost by John Milton." *Salem Press Encyclopedia Of Literature* (January 2014): *Research Starters*, EBSCO*host* (accessed April 12, 2015).

the Romantics, who saw themselves as Prometheus from Greek mythology or Satan from *Paradise Lost*, rebelling against the old ideas that would hold them and humanity back.[17]

The Romantics, including Shelley's circle, questioned religion and frequently identified as atheists, but they were not the only ones struggling with religious questions in society.[18] The fast pace of technological and philosophical change forced people to reconsider their preconceived notions, including religion. This created a level of religious turmoil that allowed Shelley and others to write texts that reflect the unique state of religion at the time of its creation. One popular concern at the time that *Frankenstein* confronts directly is the worry that that rapid technological advances would lead to humans attempting to usurp God's creative power.[19] This concern was relevant to the Romantics, who despite their focus on nature, hoped to achieve a near-divine level of creation through human effort.[20] The tensions between this fear and desire play out throughout the book and in the conflicted way the characters, particularly the Creature and Frankenstein deal with their own existence.

It is because of this tension that religion is explicitly woven into *Frankenstein* from the very beginning. The epigraph comes from *Paradise Lost*, in which Adam asks, "Did I request thee, Maker, from my clay / To mould me man? Did I solicit thee / From darkness to promote

---

[17] David Ketterer, *Frankenstein's Creation* (Victoria, Canada: English Literary Studies at the University of Victoria, 1979), 19. It is also relevant to note here that the full title of Shelley's novel is *Frankenstein; or, The Modern Prometheus*.

[18] Her father, William Godwin, began his career as a radical Protestant but had become an atheist by the time his daughter was born. Even before he abandoned religion, William Godwin used his writings to criticize the institutions of his day, including the religious establishment. Percy Bysshe Shelley, Mary Shelley's husband, was expelled from Oxford for writing a pamphlet on atheism. These two men were among the most profound influences on Mary Shelley's life, and their religious beliefs reflect the kind of environment in which Shelley was living. "Godwin, William." *Funk & Wagnalls New World Encyclopedia* (2014): 1p. 1. *Funk & Wagnalls New World Encyclopedia*, EBSCO*host*(accessed April 19, 2015). Lucy Morrison and Staci L. Stone, *A Mary Shelley Encyclopedia* (Westport, Connecticut: Greenwood Press, 2003), 173, 410.

[19] *Encyclopedia of Religion*, s.v. "Fiction: The Western Novel and Religion," by Lindsay Jones, accessed 4 April 2015.

[20] George Levine, "The Ambiguous Heritage of *Frankenstein*" in *The Endurance of Frankenstein* ed. George Levine and UC Knoepflmacher, 3-30, (Berkeley, California: University of California Press, 1979), 9.

me?"[21] This draws an immediate parallel between the creation of Adam by God in *Paradise Lost*, and by extension the Bible, and the creation of the Creature by the scientist Victor Frankenstein.[22] This comparison has resonances on both sides of the tension, but it is largely possible due to Shelley's own atheism and her sense of religion as intellectually stifling.[23] Her own indifference towards religion made it possible for her to consider a scenario in which the ultimate creative power of life is bestowed on a mere human through scientific advances, a thought which in other circumstances would be considered to blasphemous to publish.[24]

Like God in Genesis, Frankenstein labors and carefully creates his masterpiece. The scientist narrates the experience, "[h]ow can I describe my emotions at this catastrophe, or how delineate the wretch whom with such infinite pains and care I had endeavored to form? His limbs were in proportion, and I had selected his features as beautiful."[25] When God first sees His creations, the event is described by Raphael in *Paradise Lost* as "[h]ere finish'd he, and all that he had made/View'd, and behold all was entirely good."[26] Frankenstein, on the contrary, runs from his new creation, calling it "miserable monster" and "demoniacal corpse."[27]This contrast is made explicit by the Creature's own words to his creator on the subject of Adam in *Paradise Lost*: "his state was far different from mine in every other respect. He had come from the hands of God a perfect Creature, happy and prosperous, guarded by the especial care of his Creator."[28]

---

[21] Mary Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 3. Quoting John Milton, *Paradise Lost* (New York: Barnes and Noble, 2004), 10.743-745.
[22] Lucy Morrison and Staci L. Stone, *A Mary Shelley Encyclopedia* (Westport, Connecticut: Greenwood Press, 2003), 322.
[23] Schiefelbein, Michael. "'The Lessons of True Religion': Mary Shelley's Tribute to Catholicism in 'Valperga'." *Religion & Literature*, 1998., 59, *JSTOR Journals*, EBSCO*host* (accessed February 1, 2015).
[24] George Levine, "The Ambiguous Heritage of *Frankenstein*" in *The Endurance of Frankenstein* ed. George Levine and UC Knoepflmacher, 3-30, (Berkeley, California: University of California Press, 1979), 6.
[25] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 35.
[26] John Milton, *Paradise Lost* (New York: Barnes and Noble, 2004), 7.529-531,548-549.
[27] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 36.
[28] Ibid., 90.

This statement seems to align with the religious conception that science is moving towards dangerous territory by attempting to mimic God.

The novel is frequently interpreted as saying that no matter how hard Frankenstein tries, he is a flawed human and therefore unable to create anything truly beautiful or good since he creates it in defiance of God.[29] This is how Mary Shelley framed her own understanding of the story, at least in public, describing the monster in her initial dream as "[f]rightful..., for supremely frightful would be the effect of any human endeavor to mock the stupendous mechanism of the Creator of the world."[30] This is also supported by Frankenstein's own conception of himself in relation to his creation and the universal order. The Creature's identification with both Adam and Satan invites the reader to think of Frankenstein as God.[31] In the end, however, Frankenstein considers himself more like the rebellious angel, telling Walton, "like the archangel who aspired to omnipotence, I am chained to eternal hell."[32] This suggests that Frankenstein considers his sin to have been that of Lucifer: trying to usurp the position of God.[33]

However, the text does not support this reading as the only authoritative one. Instead, it is possible to argue that the difference between God and Frankenstein is not in their creative ability, but in their reaction to their creations. God declares his creation good, while Frankenstein rejects the Creature at the instant of his "birth," regardless of his potential. It is the scientist's repulsion

---

[29] *Encyclopedia of Religion*, s.v. "Fiction: The Western Novel and Religion," by Lindsay Jones, accessed 4 April 2015.
[30] Mary Shelley, *Frankenstein: or, The Modern Prometheus* (New York: Collier Books, 1961), 10, quoted in Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 36.
[31] George Levine, "The Ambiguous Heritage of *Frankenstein*" in *The Endurance of Frankenstein* ed. George Levine and UC Knoepflmacher, 3-30, (Berkeley, California: University of California Press, 1979), 7.
[32] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 152.
[33] *Encyclopedia of Religion*, s.v. "Fiction: The Western Novel and Religion," by Lindsay Jones, accessed 4 April 2015.

towards and rejection of his Creature that leads to the consequences that haunt him for the rest of the novel, not the act of creation itself.[34]

This contrast of creation and rejection shapes the Creature's self image and his relationship to Frankenstein. Over the course of the novel, the Creature reads *Paradise Lost* and initially he says, "I often referred the several situations, as their similarity struck me, to my own. Like Adam, I was apparently united by no link to any other being in existence."[35] However, over the course of the novel, his perspective and identification change.[36] He reprimands Frankenstein:

> Accursed creator! Why did you form a monster so hideous that even YOU turned from me in disgust? God, in pity, made man beautiful and alluring, after his own image; but my form is a filthy type of yours, more horrid even from the very resemblance. Satan had his companions, fellow devils to admire and encourage him, but I am solitary and abhorred.[37]

The Creature takes issue with Frankenstein for his abandonment, but also for his solitude, which God remedied by creating Eve, but Frankenstein refuses to do for the Creature, out of fear and loathing of his creation.

In *Frankenstein*, Mary Shelley took the classic creation myth of Genesis and adapted it to reflect the tensions present in her own tumultuous society. The contrasting reactions to the massive social changes of this period are encapsulated in the different possible readings of the novel, which address both the secular doubt about the role of God in the ever-changing world and the religious concern about the dangers of a world where humans take on the responsibilities

---

[34] Jasia Reichardt. "Artificial Life and the Myth of Frankenstein" in *Frankenstein Creation and Monstrosity*, ed. Stephen Bann, 135-157 (London: Reaktion Books, 1994), 137.
[35] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 90.
[36] Elsie B. Michie, "Marx's Theories of Alienated Labor" in *Approaches to Teaching Mary Shelley's Frankenstein*, ed. Stephen C. Behrendt 93-98 (New York: The Modern Language Association of America, 1990).
[37] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 91.

of God.[38] The rapid scientific advancements and the precarious and unsettled world of Mary Shelley provided the perfect place for the creation of the genre of science fiction, which raises questions about the future and how technology functions in society in ways that had never been necessary before.

The fascination with created beings continued into the twentieth century, and another important shift occurred with the advent of the assembly line and mass production. These advances fundamentally changed fictional creation of automata. Instead of a single scientist animating a single corpse or a single rabbi creating a single golem, the creation of automata became an industrial and technological enterprise on a much larger scale. The word "robot" was coined by the Czech writer Karel Čapek in his 1921 play *R.U.R*. The title is the abbreviation for Rossum's Universal Robots, the corporation that creates robots in Čapek's universe, and around which the action of the play takes place. Čapek derived "robot" from the Czech root of "slave" or "worker."[39] Čapek's robots, are not, however, the mechanical beings that contemporary society associates with that word.

These first "robots" are made from a previously unknown organic substance and mass-produced in an isolated factory. Their creators tightly control their emotions and abilities. In act one, for instance, "the Robots feel practically no bodily pain," but the scientists are developing pain-nerves for the Robots. When asked why they would do such a thing, the scientist responds, "[f]or industrial reasons… Sometimes a Robot does damage to himself because it doesn't hurt him. He puts his hand into the machine, breaks his finger, smashes his head, it's all the same to

[38] Justo Gonzalez, *A History of Christian Thought, Volume 3*, 2nd ed (Nashville, Tennessee: Abingdon Press,1987), 335. George Levine, "The Ambiguous Heritage of *Frankenstein*" in *The Endurance of Frankenstein* ed. George Levine and UC Knoepflmacher, 3-30, (Berkeley, California: University of California Press, 1979), 10.
[39] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 164.

him. We must provide them with pain. That's an automatic protection against damage.[40] This gradual progression to Robots with more human-like abilities is a theme that recurs in later stories of artificial intelligence, though the actual artificial beings in *R.U.R.* are very different from the modern conception of robots.

R.U.R represents an important step in the development of the modern concept of the robot or artificial intelligence. While Shelley introduced scientific principles into artificial creation, Čapek brought the artificial creation into the 20th century by mechanizing it and making the process of gaining intelligence and personhood a gradual one. As valuable as these contributions are, however, the contemporary notion of AI could not fully develop until the invention of the computer.[41] In order develop a created intelligence in a scientifically realistic way, there first had to be programmable machines that could house that intelligence.

In 1936, Alan Turing published "On Computable Numbers," which is considered to be the first paper in the field of computer science. Turing, a British mathematician and logician who lived from 1912 to 1954, is recognized as the founder of the general field of computer science, as well as the specific subfield now known as artificial intelligence. During World War II, Turing made a non-programmable precursor to the computer in order to break the code of the German Enigma Machine. After the war, Turing continued his work in the field of computer science, and helped to create some of the earliest computers, while also developing theories of computer science and machine thought that remain relevant in the field.[42]

---

[40] Karel Čapek, *R.U.R.*, trans. Paul Selver and Nigel Playfair (New York: Doubleday, 1923).
[41] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 162.
[42] Stuart M. Shieber, introduction to *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, ed. Stuart M. Shieber (Cambridge, Mass.: MIT Press, 2004), 5.

Turing himself admitted that it took the invention of the computer to bring about a renewed interest in artificial beings or "thinking machines."[43] In 1950, Turing published the paper "Computing Machinery and Intelligence" in *Mind* magazine, which laid the foundation for much of the current study of AI. In this paper, he proposed the now-famous Turing test for determining the intelligence of a machine. Herzfeld explains that "[t]wo questions stand at the heart of the AI endeavor. What is intelligence? And second, how would we know if a computer possessed intelligence? There is no simple answer to either of these questions."[44] Recognizing even early on that empirically measuring the intelligence or consciousness of a machine is impossible, Turing developed a definition of intelligence similar to Justice Potter Stewart's definition of obscenity as "I know it when I see it." He reasoned that the best way to determine the intelligence of an artificial being is to test it in conversation with a human, a being we already consider to be intelligent.[45]

The Turing Test, which Turing himself referred to as "The Imitation Game," consisted of three actors: two humans and the machine being tested. One of the humans would act as the interviewer, having an individual conversation with the human and the machine in turn. These conversations would take place in such a way that the human interviewer could not see the two test subjects, but only receive their written responses. Turing set no parameters for the topics of conversation or the behavior of the interviewer, only that the human was to engage in a conversation with each of the two other actors, asking questions and receiving answers. If the human interviewer is unable to distinguish between the machine and the human after repeated tests, then the machine passes, and is to be considered a thinking machine. Turing did not believe

---

[43] Ibid., 61.
[44] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 33.
[45] Stuart M. Shieber, introduction to *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, ed. Stuart M. Shieber (Cambridge, Mass.: MIT Press, 2004), 5.

that it would be possible to prove the "intelligence" of a machine, but instead believed that if a machine were able to perform the functions of intelligence according to the standards of another intelligent being (in this case a human), then it deserved to be considered an intelligent.[46] This is criteria seems possible, as Herzfeld explains, because "[d]iscourse is unique among human activities in that it subsumes all other activities within itself, at one remove. If we accept the Turing Test, as most of the AI community has, as the ultimate arbiter of intelligence, then we have defined intelligence relationally."[47]

This is far from the only paradigm for determining the intelligence of a machine, but since it is considered the founding text of the field of artificial intelligence, it is by far the most frequently discussed and influential. Philosopher John Searle formulated the most important refutation of this theory. He presented his counterargument in the form of the Chinese Room thought experiment. Searle compared a Turing-passing machine to a room with a person sitting inside. The person inside this room does not speak Chinese, but has a book full of Chinese characters and symbols. The person in the room is then passed pieces of paper containing Chinese characters on them. The man then follows the rules presented by the book, which tell him what to do with the symbols, and he then sends a slip of paper out of the room with Chinese characters on it. If the person in the room learned to perform this operation sufficiently quickly, he could convince someone outside the room, who only knew the inputs and outputs that he spoke Chinese. This would not be the case, since the person in the room does not understand what he is communicating; he is only following preordained rules. Searle uses this thought experiment to differentiate between a being that is actually conscious, thinking, and able to

---

[46] Alan M. Turing, "Computing Machinery and Intelligence" in *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, ed. Stuart M. Shieber (Cambridge, Mass.: MIT Press, 2004) 68.
[47] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 46.

understand and communicate, and one who is just a sufficiently advanced mimic. This distinction remains relevant in the realm of artificial intelligence even today, with most attempts at passing the Turing test being more similar to the Chinese Room than an actual thinking being.[48]

It was not until 1956, however, that the field Turing founded was given a name. John McCarthy, an important early researched in the field of AI coined the term for a conference he was giving on this sub-field of computer science.[49] This name incorporates Turing's idea of the "intelligent" machine, as one whose output indicates some kind of mind or thought process to a human observer. The phrase also carries a connotation of non-organic life, unlike a Frankenstein's monster or earlier organic automata. It is an "artificial" intelligence since it is mechanical and not of the natural world, but was instead created by humans.[50] This terminology of artificial intelligence has become an important part of conversations across many disciplines to describe and understand the human attempt to create new beings.

As the idea of robots and the precursors to artificial intelligence were beginning to permeate the culture, Isaac Asimov wrote his first Robot stories. The most influential stories were written and published as stand-alone stories in science fiction magazines, including John W. Campbell Jr.'s *Astounding Science Fiction* between 1940 and 1950. These stories that share characters and themes, were then put together with a frame narrative and collected in the book *I, Robot*, published in 1950. The idea of mechanization and industrialization that had been so central to Čapek's story had become commonplace, and the rapid advancement of other machinery suggested a future in which intelligent and humanoid robots could be possible. Warrick claims that Asimov is "deservedly regarded as the father of robot stories in SF." This is

[48] Niran B. Abbas, *Thinking Machines: Discourses of Artificial Intelligence* (Hamburg: Lit, 2006), 87-89.
[49] "Artificial Intelligence" *Funk & Wagnalls New World Encyclopedia* (2015): 1p. 1.
[50] Niran B. Abbas, *Thinking Machines: Discourses of Artificial Intelligence* (Hamburg: Lit, 2006), 11.

partially due to the fact that only "[t]hree stories using electronically operated robots appeared before Asimov's first story, 'Robbie' was published in 1940."[51] Asimov's stories became an important part of the conversation about robots and artificial intelligence. They postulate the advancement and development of a robot or AI based on the processes of scientific development. These stories also highly influenced the literature and thought surrounding these theoretical beings and contributed to the desire of many later computer scientists to implement the ideas that Asimov posited in his stories.

Asimov continued to write stories in the same universe as *I, Robot* for the rest of his life, and explored the concept of artificial intelligence in other ways, but the original nine short stories had the biggest impact.[52] In these Asimov coined both the terms "positronic brain" and "robotics" as a field of study and inquiry.[53] However, Asimov's most important contribution to the field of robotics and of science fiction in *I, Robot* is his creation of the Three Laws of Robotics. The Laws state:

1) A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2) A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law.
3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.[54]

These Laws are the foundation of any positronic brain in the *I, Robot* universe. One of the characters remarks, "I needn't tell you, Dr. Calvin, that there has always been strong opposition

---

[51] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 54.
[52] Of Asimov's other Robot stories, probably the most important is *The Bicentennial Man*, a 1966 Hugo-winning novelette about a robot who wants to become human. However, Asimov also wrote a series of five novels, beginning with *The Caves of Steel*, focusing on the human detective Elijah Baley and his robot partner R. Daneel Olivaw, as well as many more short stories that focus on similar themes to the initial short stories.
[53] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 165-166.
[54] Isaac Asimov and Janet Asimov, *It's Been a Good Life* (Amherst, N.Y.: Prometheus Books, 2002) 61. Over the course of his life Isaac Asimov wrote three volumes of autobiography: *In Memory Yet Green* (1979), *In Joy Still Felt* (1980), and *I, Asimov* (1994). After his death, Asimov's wife, Janet, edited these three volumes into *It's Been a Good Life*, a single autobiography spanning the entirety of Asimov's life.

to robots on the Planet. The only defense the government has had against the Fundamentalist radicals in this matter was the fact that the robots are always built with an unbreakable First Law---which makes it impossible for them to harm human beings."[55] Despite this resistance, humans in this world usually accept robots among them, but only because they know that no robot can be created without or disobey the Laws.[56]

One of the reasons that Asimov came up with these rules is to combat what he calls the "Frankenstein Complex." Named after the character in Mary Shelley's novel, Asimov first uses this term in his 1947 story "Little Lost Robot." In addition to inventing the genre, Shelley's nuanced and fully realized vision of an artificial being had a profound effect on the portrayal of robots and other artificial creations in later science fiction. Asimov even termed the tendency of science fiction writers after Shelley to tell stories of human creations seeking to dominate or destroy their creators "the Frankenstein complex." It refers to the fear of machines destroying their creators or taking over humanity.[57] Asimov coined this term out of frustration at the then-ubiquitous trope of a robot turning on its human masters. As Brian Stableford claims in *The Sociology of Science Fiction*, "Asimov… was determined to put across the point that the weakness giving rise to the anxiety was in man, and that the stigmatization of the machine was unjustified"[58] As an author who took the science part of science fiction writing very seriously, this frustrated Asimov. He did not find the idea that humans would create robots who could destroy them compelling.

---

[55] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 115

[56] Several stories in the collection, most notably "Reason" and "Little Lost Robot,", deal with the implications of robots disobeying, misinterpreting, or being created with exceptions to these Laws. However, in Asimov's original stories, this behavior is almost always resolved without harm to humans and both the conflict and resolution are situated firmly within the context of the Three Laws. In addition, this type of plot is a far less common and far less catastrophic in Asimov's stories than it is in many science fiction stories with a similar premise, including the film adaptation in-name-only of *I, Robot*.

[57] Jeff Puncher, *Brave New Words: The Oxford Dictionary of Science Fiction* (Oxford: Oxford University Press, 2007), 67.

[58] Brian M Stableford, *The Sociology of Science Fiction*. (San Bernardino, Ca.: Borgo Press, 1987), 107.

Instead of fearing them, Asimov argued that robots were just another kind of technology, a tool created by humans for human use. Just as humans have constantly worked to improve the safety of tools and technologies, from knives to cars, so that they do not hurt people, the creators of the robots would take all reasonable precaution to prevent them from hurting humans. Asimov was in accordance with "[t]he dominant opinion within the genre---often actively didactic---[which] held that this anxiety was unjustified and that the fear was misdirected. The real danger, according to the science fiction stories of the day, was not the machines themselves, but our moral and intellectual inability to deal with them."[59] Asimov took this to its logical conclusion, and therefore created a system in which it was impossible for this technology to be used to harm humans or rise up against them.[60]

These Laws fundamentally changed the ways people discussed and viewed robots and artificial intelligence. Asimov himself immodestly but accurately suggests that since he introduced these Laws in his story "Runaround" in 1942, every author who discusses robots, in some way, has had to address the Laws. Many authors essentially choose to use the Laws by another name or without explicitly stating them,[61] while others choose to have robots that violate these Laws, but must justify their absence.[62] Asimov believed that the Frankenstein complex plot limited the potential for robot narratives in science fiction, and in his Three Laws, though they are rarely quoted by name, he provided authors with a way to create and justify new kinds of robot narratives.

---

[59] Ibid., 110.
[60] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 165.
[61] The television show *Humans*, a 2015 joint production of the British Channel Four and the American AMC, for example, refers to the "Asimov controls" as part of the programming of the artificial intelligences in that universe that means that they cannot harm a human, nor does their programming allow them to end their own existence, as is consistent with the Third Law.
[62] Isaac Asimov and Janet Asimov, *It's been a Good Life* (Amherst, N.Y.: Prometheus Books, 2002) 62.

This extends not only to authors, but also to those actually attempting to build intelligent robots. Asimov mentions conversations with real-life roboticists and AI researchers, including Marvin Minsky, an important early figure in the field of artificial intelligence, who were inspired by his Robot stories. Warrick tells the story of "Joseph Engelberger, builder of the first industrial robot, Unimate (1958), [who] attributes his long-standing fascination with robots to his reading Asimov's *I, Robot* when he was a teenager."[63] In Asimov's experience, these scientists take the theory of artificial intelligence proposed in the Three Laws seriously and try to incorporate them into their practical work on thinking machine.[64] This influence continues to the present day, as evidenced by paper published in 2007 by Lee McCauley, wherein he explores both why the Laws as written are impossible to implement and why the AI community must nonetheless try to uphold the spirit of these Laws.[65]

In Asimov's life it is possible to identify two main factors that led to the creation of the Laws and their narrative, as well as in-universe purpose. The first is his scientific background. Asimov received a doctorate in Biochemistry from Columbia University in 1948, and worked at the Boston University Medical School until 1958, when he decided to devote himself to writing full-time. He published numerous books on science for both popular and academic audiences, and was considered by fellow science-fiction writer Arthur C. Clarke to be both the best science writer and the second-best science fiction writer in the world.[66] This familiarity with the scientific method and lifelong immersion and interest in science and technology ensured that he

---

[63] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 34.
[64] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 167.
[65] Lee McCauley, "The Frankenstein Complex and Asimov's Three Laws" in *Association for the Advancement of Artificial Intelligence*, Published May 2007, https://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-003.pdf
[66] This was agreed upon in a "meeting" held between the two authors in the back of a taxi, and the accord became referred to as the "Clarke-Asimov Treaty of Park Avenue." Under its stipulations Asimov agreed that Clarke was the best science fiction writer and the second-best science writer in the world (after Asimov himself), and Clarke conceded that Asimov was the best science writer and second-best science fiction writer in the world (after Clarke himself).

had a realistic idea of how an artificial intelligence might develop and change as history progressed. It also led him to the logical conclusion that humans would take the same precautions in the development of AI as they would with other technology, and his version of these safeguards is the Three Laws of Robotics.

The other major factor in the creation of Asimov's Three Laws is his theory of science fiction. In his 1953 essay "Social Science Fiction," Asimov argues that there are three types of science fiction: gadget, adventure, and social.[67] The most interesting kind of science fiction for Asimov and most other science fiction scholars is social science fiction. Asimov describes his three kinds of science fiction in terms of an "invention," which is the concept or element that exists in the fictional world but not in ours and makes the work science fiction. In social science fiction this invention is used to engage in social commentary or to explore the effect of this change on society. In this sub-genre, the plot is shaped by the new convention, but not myopically focused on its creation or technological specifications, as gadget science fiction is.

When explaining the types of science fiction using the automobile as an example, Asimov suggested that a social science fiction piece might predict traffic accidents, urban sprawl, and pollution. This shows well how difficult it is to create social science fiction, but also how important the genre can be for society.[68] According to his belief in social science fiction, it is not enough for Asimov, that an author merely discusses the development of the robot, but the author must also examine the effect of this development on society. Warrick sees this in Asimov's work, noting that his "stories are often concerned with the same themes: the political potential of the computer, the uses of computers and robots in space exploration and

[67] Asimov identified gadget science fiction as focused on the technical aspect, showing how the invention was made, and adventure science fiction as an adventure story that featured some invention as a plot device.
[68] Isaac Asimov, "Social Science Fiction" in *Modern Science Fiction, its Meaning and its Future*, ed. Reginald Bretnor and John W. Campbell Jr. (Chicago : Advent Publishers, 1979; 2d ed, 1979), 171.

development, problem solving with computers, the differences between man and machine, the evolution of artificial intelligence, the ethical use of technology."[69] If Turing's 1951 paper both asked and posited an answer to the question "how does one identify an artificial intelligence?" Asimov's Robot stories ask, "How would society create an artificial intelligence?" And "what effect would this artificial intelligence have on society?"

This distinction is echoed in the works of science fiction theorists such as Darko Suvin, and Asimov's own mentor and publisher John W. Campbell Jr. Suvin proposes a very limited definition of science fiction, but identifies it strongly with the concept of an intellectually based novum. He defines the novum as "a totalizing phenomenon or relationship deviating from the author's and implied reader's norm of reality."[70] Within science fiction, there are different ways for the author to create this novum and for the plot and characters to interact with this newness. The most common model for this relationship is referred to as extrapolation, which is taking a scientific concept and positing the effects of the further development of this concept. Campbell argues that this concept can be a hard science concept such as space travel, or a soft science concept such as increased censorship, but as long as both expand the novum along intellectual lines, they are considered to science fiction.[71]

Another way of addressing the novum is through analogy or metaphor. In this method, the direct relationship between current technology or ideas and the novum is less pronounced. Instead, a science fiction story focusing on analogy tends to be about a technology or culture farther removed from the current science and act far more as a commentary on contemporary

---

[69] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 57.
[70] Darko Suvin, *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre* (New Haven: Yale University Press, 1979), 64.
[71] John W. Campbell Jr., "The Science of Science Fiction Writing" in *Of Worlds Beyond: The Science of Science Fiction Writing*, ed.Lloyd Arthur Eshbach, (Reading, Pennsylvania: Fantasy Press, 1947), 87.

society or technology.[72] In her introduction to *The Left Hand of Darkness*, a classic of analogical science fiction, Ursula Le Guin defines this type of science fiction:

> Science fiction is metaphor. What sets it apart from older forms of fiction seems to be its use of new metaphors, drawn from certain great dominants of our contemporary life---science, all the sciences, and technology, and the relativistic and the historical outlook, among them. Space travel is one of these metaphors; so is an alternative society, an alternative biology; the future is another. The future, in fiction, is a metaphor.[73]

Asimov's Robot stories, and the Three Laws in particular, exhibit both of these types of science fiction. Though Asimov did not believe at the time he wrote the original Robot stories that artificial intelligence was a possibility, his scientific background nonetheless allowed him to carefully extrapolate the development of the robot from the creation of the positronic brain.[74] He claims that he saw robots as "machines not metaphors," and that this is the foundation of the success of his Robot stories.[75] This attitude is shown in *I, Robot*, on multiple occasions, but especially in those focusing on Powell and Donovan, whose job it is to test new kinds of robots. In "Runaround," when they are discussing the older robots whose replacements they are testing, Donovan says, "they may be subrobotic machines. Ten years is a long time as far as robot-types are concerned." Powell responds, "they're robots…. They've got positronic brains: primitive, of course."[76] This shows that even though artificial intelligence is possible in this society, the development of this technology is ongoing and older models are outdated compared to the newest versions.

Despite this strong extrapolative element, the stories in which these scientific developments occur are clearly social science fiction with a distinctly metaphorical bent. In one

---

[72] Darko Suvin, *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre* (New Haven: Yale University Press, 1979), 75.
[73] Ursula K. Le Guin, *The Left Hand of Darkness* (New York: Ace Books, 1969), xix.
[74] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995), 174.
[75] Ibid., 166.
[76] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 28.

of the most striking passages in the text of *I, Robot*, one character says, "[i]f you stop to think of it, the three Rules of Robotics are the essential guiding principles of a good many of the world's ethical systems," and goes on to explain that anyone who is following the Three Laws "may be a robot, and may simply be a very good man."[77] Asimov did not intend to write morality tales about artificial beings, but his stories nonetheless explore the difficulties of creation, exclusion, and ethics in a way that is highly applicable even to a world without robots. Asimov's mentor, John W. Campbell Jr. explains, "above all else, a story-science fiction or otherwise---is a story of human beings. If it's a thinking robot that's the hero, then the robot is made either practically human, or is aligned against human characters for whom we're rooting."[78] As much as Asimov made his robots machines, the humans bring the drama and weight at the heart of the stories.

Asimov himself was a staunch atheist throughout his life. His family was Jewish, but after they immigrated they were all but non-practicing. Asimov's first wife Gertrude was Jewish, and they were married in a Jewish ceremony, but Asimov did not consider himself to be Jewish.[79] He even came into conflict with some of the other Jewish servicemen serving at his the naval base during the war when they wanted time off for Yom Kippur, but Asimov was indifferent.[80] Especially as he grew older, Asimov grew more and more dedicated to rationality and humanism. Asimov was a scientist both by training and inclination, and he worked hard to incorporate scientific ideas of rationality into his belief system as well as his writing.

Asimov, as mentioned above, did believe that religion had some place in science fiction, and that religion must inevitably be involved in many of the questions that science fiction sought to ask, but that did not mean it must not question religion as a genre. Asimov also admitted that

---

[77] Ibid., 182.
[78] John W. Campbell Jr., "The Science of Science Fiction Writing" in *Of Worlds Beyond: The Science of Science Fiction Writing*, ed. Lloyd Arthur Eshbach, (Reading, Pennsylvania: Fantasy Press, 1947), 87.
[79] Isaac Asimov and Janet Asimov, *It's Been a Good Life* (Amherst, N.Y.: Prometheus Books, 2002), 71, 19.
[80] Ibid., 74.

despite his Jewish heritage, the majority of the religion that wound up in his stories is Christianity, since that was the dominant religion that had been around for his life, so that was what he was familiar with. Asimov also wrote several books later in life about religion, including *Isaac Asimov's Guide to the Bible* and *In the Beginning: Science Faces God in the Book of Genesis*.[81]

The later theorist Darko Suvin, for whom religion, myth, and even fairy story and fantasy are all inimical to true science fiction, takes Asimov's emphasis on the secular to an extreme. He argues that "[t]he myth is diametrically opposed to the cognitive approach since it conceives human relations as supernaturally determined."[82] Suvin, looking back on the genre in 1979, strictly limits the definition of what can be science fiction in a way that excludes all religious impulse, reference, or inspiration.

In the decades since Suvin and Asimov, attitudes towards religion and science fiction have shifted dramatically, as have the emphasis of the genre and the makeup of the authors. However, even in this time, it is possible to see that as much as Asimov and Suvin prided themselves on rationality and empiricism, Asimov was more correct when he claimed that "it is the very essence of literature that it consider the great ideas and concerns of human history. Surely the complex of ideas that goes under the head of "religion" is one of the most central and essential, and it would be rather a shame to have it declared out of bounds."[83] Even Suvin concedes that, like religion, "this genre raises basic philosophical issues" about the destiny of humanity, the consequences of human action, the definition of a person, and the nature of being.

---

[81] Ibid., 305.
[82] Darko Suvin, *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre* (New Haven: Yale University Press, 1979), 7.
[83] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995), 230.

Isaac Asimov's *I, Robot* stories are an important turning point in the field of artificial intelligence, science fiction, and even the way people talk about robots and technological advancement in casual conversation. His combined scientific and literary experience allowed him to ask questions of this burgeoning field in ways that resonated with scientists and authors alike. His work raised issues not only of how to write and think about robots, but also of how to build and program robots in the real world. There are many technical questions that the AI field is still seeking to solve, but the ideas that Asimov presented in his first short stories still guide the direction of development and the ethical conversations within AI. After the initial nine short stories Asimov himself continued to write and think more about robots. Over the course of his life he wrote more than thirty short stories and five novels about robots, in addition to several essays later in life about the scientific advancements in robotics.[84]

The questions Isaac Asimov asked and the concepts he invented in these stories remain at the heart of much thought about AI in historical reflection, present application, and future speculation. As Michael Pinsky claims, "Science fiction, which has throughout its history as a genre been treated most often as a sort of literary ghetto: a marginalized (critically suspect) literature about marginalized voices (alien and alienated others) by frequently marginalized authors (successful only by the standards of a fringe fan culture)."[85] Asimov's thoughts on of obedience, emotion, ethics, creation, and what it means to be human are as relevant to contemporary culture as they were seventy-five years ago. Indeed, they are perhaps even more important now, as scientists that Asimov inspired begin to bring the beings he imagined closer to reality. As Herzfeld argues, "Whether computers, our 'mind children,' as Moravec calls them, are positioned to replace humanity or coexist with us, whether we even wish to pursue the dream

[84] Ibid., 168.
[85] Michael Pinsky, *Future Present: Ethics And/As Science Fiction*, (Madison, NJ: Fairleigh Dickinson UP, 2003), 15.

of AI at all, depends on which aspect or aspects of our own nature we hope to copy in our attempt to create autonomous machines."[86]

[86] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), ix.

CHAPTER 2

The Question of Personhood

Asimov claimed that the robots in his stories were effective since they were "machines not metaphors." [87] Warrick and Herzfeld likewise try to impose a dichotomy on AI stories, with Warrick suggesting that "writing about artificial intelligence requires dealing with a mechanical form that is… logical [and] mathematical…causes and effects and relationships are fixed. Such a mechanistic, closed model is anathema to the creative mind, which tends to work by intuition and not logic."[88] Despite this common view, as Campbell points out, "above all else, a story---science fiction or otherwise---is a story of human beings. Even if a dog is the central character, we're actually projecting human qualities into that central character, and watching only the human-like characteristics of the dog."[89] From the beginning of artificial intelligence narratives with *Frankenstein*, there has been an understanding that the act of creating another being is a complicated and dangerous one. While Dr. Frankenstein set out to create a person, Asimov did not, and his characters have complex relationships to the concept of personhood. [90] Warrick explains that Asimov asks, ""What is a man? What is a machine? These questions intrigue Asimov... But what happens to man's image of himself when machines begin to acquire some of these characteristics? If machine intelligence can perform the functions of human intelligence, is

---

[87] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995), 164.
[88] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 6.
[89] John W. Campbell Jr., "The Science of Science Fiction Writing" in *Of Worlds Beyond: The Science of Science Fiction Writing*, ed. Lloyd Arthur Eshbach, (Reading, Pennsylvania: Fantasy Press, 1947), 87.
[90] For the purposes of this discussion, a person will be considered any conscious and intelligent being worthy of rights, while a human will refer specifically to a being classified as a homo sapiens

man nothing more than a machine?"[91] Throughout the *I, Robot* stories, Asimov examines both humans and robots that possess various degrees of personhood and person-like behavior in an attempt to explore some of these questions.

Since the stories move more or less chronologically and the machines change throughout the stories, it is interesting to examine the different ways in which the robots across time are and are not portrayed as persons. A valuable metric for this is the Turing Test. As proven by Searle, it is an imperfect metric, but it is a convenient starting standard for examining how Asimov's robots relate to ideas of creation, consciousness, and personhood. Aside from Turing's definition, another useful paradigm of consciousness is Warrick's understanding that "[a]s the child develops, his image grows. He perceives himself as separate in the midst of a world of objects. Consciousness or self-awareness has begun."[92] Essentially Warrick equates consciousness with self-awareness, which is presumed to require some level of intelligence and independent identity. While both of these standards are imperfect to understand a concept as inscrutable as consciousness, they both provide useful frameworks with which to examine the potentially conscious beings of Asimov's stories.

The first robot to whom the reader is introduced is Robbie, a robot nanny from the story of the same name. Robbie could not pass the Turing Test since, as Dr. Susan Calvin explains, "Robbie was a non-vocal robot. "[93] Though his positronic brain operates under the same basic principles as do those of the more advanced robots in later stories, Robbie's earlier model is limited in many ways. [94] However, he does seem to possess an independent sense of self, and

---

[91] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 62.
[92] Ibid., 7.
[93] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), xv.
[94] Though gender is not a major factor in the ways the robots interact with humans in the story, all the robots are given names, frequently male names (Herbie, Robbie, Dave), but are always referred to as "he," "him," or (rarely) "it" in the third person, so I have followed that convention.

Herzfeld argues that when thinking of intelligence, "we must consider such abilities as movement, speech, and contemplation of the world, indeed, awareness of the self within the world."[95] Despite his technical failure of the Turing Test, Robbie is portrayed as far more of a person than many of the robots presented in later stories. This makes more sense given the fact that the Turing Test is not actually primarily a test of verbal skills.

As Herzfeld acknowledges, "Turing considers the ability to relate in discourse, as human beings do, to be far more important than accuracy or precise functioning in any realm.... Turing notes here that intelligence goes far deeper than mere competence."[96] Fundamentally, Turing does not suggest that intelligence is about speech, but rather about communication and connection. It is in this way that Robbie is most a person. He experiences, or at least exhibits emotions. He is described as "hurt at the unjust accusation," and he finds "Gloria's mother… a source of uneasiness."[97] He is able to request that Gloria tell Cinderella over any other story, and he at least gives the appearance of valuing Gloria's stories and spending time with her.

The interplay of the Laws and the ways in which these robots interact with each other and with humans are so complex that Asimov had to invent a new science in order to examine them. Susan Calvin, a recurring character and the frame narrator, is a robopsychologist, a term coined by Asimov to describe the new field of people whose job it is to understand how robots think, just as a psychologist does with people. Calvin's job is to "study the robot itself and work backward. Try to find out how he ticks." [98] Throughout the stories, she talks to robots and talks to people about robots in order to discern their motivations, predict their future behaviors, and understand their experiences.

---

[95] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 34.
[96] Ibid., 46.
[97] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 4, 6.
[98] Ibid., 114.

In "Robbie," Calvin, having only met Gloria briefly and Robbie never, does not provide a great deal of information about whether or not Robbie is actually thinking and feeling or just acting out programming to be caring and protective. The text itself contains evidence for both interpretations. Calvin's initial description of Robbie as a robot from "the days before extreme specialization, so he was sold as a nursemaid," suggests that he has adapted to the specific task of nursemaid, but that was not his initial purpose.[99] In contrast, Gloria's father analyzes Robbie as "constructed for only one purpose really---to be the companion of a little child. His entire 'mentality' has been created for this purpose. He just can't help being faithful and loving and kind."[100] This would indicate that Robbie's kindness and emotions and care for Gloria are programmed responses, not choices he himself made. The question of how Robbie's behavior is determined and how much autonomy he has in this regard is never fully resolved in any of the stories, but the different characters' perceptions of his autonomy are responsible for their varied reactions to him.

The main conflict of this story is to what extent Robbie deserves to be granted personhood. Gloria's mother, Mrs. Weston, vehemently denies Robbie any semblance of personhood and believes that is dangerous or at least unhealthy for her daughter to have such a connection to a non-human individual. She tells her husband, "I won't have my daughter entrusted to a machine---and I don't care how clever it is, it has no soul, and no one knows what it may be thinking. A child must isn't *made* to be guarded by a thing of metal."[101] Mrs. Weston protests so much that Robbie is sold back to the company, and instead of another human, Gloria is given a dog. It is interesting that Gloria's mother considers a dog a more fit companion for her child than a robot. This betrays "a prejudice against robots which is quite unreasoning," and an

[99] Ibid., xv.
[100] Ibid., 9.
[101] Ibid., 9.

irrationally strong preference for the biological over the intelligent.[102] Despite Mrs. Weston's actions, the narrative seems to argue that Robbie is closer to a person than the biological dog, and Gloria certainly thinks so. She tells her mother, "'I don't want a nasty dog---I want Robbie,'" and goes on to refute her mother's assertion that Robbie is "only a machine, just a nasty old machine" by saying, "'[h]e was a *person* just like you and me and he was my *friend*.'"[103]

The narrative supports Gloria's understanding of her nursemaid by introducing Robbie playing hide and seek with Gloria and giving insights into his view of the world, even including some emotional language. As a human Gloria had some responsibility for Robbie, since he is obligated to follow her orders, which gives her a level authority even over her nursemaid, just as she might be able to command a dog. It is important to note that the narrative seems to allow Robbie personhood, as much or more so than non-human organic beings. The question of personhood is a complicated one, as Herzfeld explains, "[a]t the root of our fascination with creating an artificial intelligence in our own image lies a continuing problem of defining what it is we wish to image---in other words, what it means to be truly human."[104]

As much as the narrative seems to argue for Robbie's personhood, and indeed the personhood of the other machines in the story, the text almost never challenges the status of the robots. Like Robbie, many of them seem to have emotions and express opinions, and others even lie or play practical jokes. This is all framed in terms of their programming, but even Asimov admitted later that the robots were fundamentally conscious. This raises the question of why it is then acceptable for the Westons to sell Robbie back to the factory. Calvin even ends the story by

---

[102] Ibid., 238.
[103] Ibid., 14.
[104] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 6.

telling the interviewer that "'Gloria had to give up Robbie eventually,'" because "'the mobile speaking robot… made all non-speaking models out of date,'" and "'most of the world's governments banned robot use on Earth.'"[105] The story, despite its humanization of Robbie and siding with Gloria and Mr. Weston over Mrs. Weston, nonetheless sees Robbie as a machine that will inevitably become obsolete, and in another story, the reader is informed that all robots "are dismantled. And the positronic brains re-used or destroyed."[106] This seems, upon further reflection, quite an inhumane way to deal with a being that seems to be conscious and capable, but this idea is rarely taken up in a significant way by the story.

There are two different kinds of robots in "Runaround," the second story in the collection, but neither of them comes anywhere near to the level of humanity exhibited by Robbie. They can speak, but they do not express emotions or seem nearly as involved with humans as Robbie. This is likely because of the change of setting and the different functions of these robots. Gregory Powell and Michael Donovan, two humans who work for US Robotics and play an important role in several of the stories are "sent… out to Mercury to help build the mining station there" with some robotic assistance.[107] This industrial setting is very different from the domestic scope of the first story, and these robots are more specialized and functional than Robbie. The main conflict of the piece does not deal with the personhood of the robots or the perception of the robots by humans; it is instead a story of the conflicts between the Three Laws.

The primary robot in the story, SPD 13, nicknamed Speedy is a "new-type robot, still experimental."[108] The plot begins when Powell and Donovan send Speedy out to get selenium,

---

[105] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 28.
[106] Ibid., 211.
[107] Ibid., 28.
[108] Ibid., 29.

which necessary to operate "the photo-cell banks that alone stood between the full power of Mercury's monstrous sun and themselves."[109] However, instead of retrieving the selenium as ordered, Speedy begins is circling the selenium pool for no apparent reason, and he is not stopping. When the humans find Speedy, he seems to be drunk. However, as Powell explains, "Speedy isn't drunk---not in the human sense---because he's a robot, and robots don't get drunk. However, there's *something* wrong with him which is the robotic equivalent of drunkenness."[110] This malfunction is even affecting his ability to obey the laws, since he ignores a direct order from Powell in violation of the Second Law.

The humans eventually discover that this behavior is a result of an unsolvable conflict between the various Laws of Robotics. This is reminiscent of the stories of computers or robots who are destroyed or confused by human emotion or conflict, a storyline which often displays the superiority of humans by their ability to reason their way through conflicts and deal with difficult and emotional situations. This trope is explored in "Catch that Rabbit," "Liar!" and "Escape," in addition to "Runaround," but the situation is generally more complicated than the robot just shutting down. In all except "Liar!" the robot responds by acting abnormally or seeming to go "crazy," which signals to the humans that something is wrong, and they can then fix it, rather than the whole robot to become unable to function. This shows that the positronic brain allows for more flexibility than is typically associated with a robot that cannot comprehend conflict or emotion.

The logical nature of Powell and Donovan's solution demonstrates Asimov's level of dedication to extrapolative robots. The robot acting drunk is not doing so because of any traditionally human reason, but because there is a conflict between the most basic information of

[109] Ibid., 32.
[110] Ibid., 43.

his programming. The selenium pool will kill him if he goes near it, so according to the Third Law, he must not go near it. Powell explains that any "conflict between the various rules is ironed out by the different positronic potentials in the brain."[111] In this case "Rule 3 has been strengthened" because "Speedy is one of the latest models… not a thing to be lightly destroyed."[112] This causes a problem because when Donovan sends Speedy to the selenium pool, he "gave him the order causally and without any special emphasis, so that the Rule 2 potential set-up was rather weak."[113] This results in "an equilibrium… Rule 3 drives him back and Rule 2 drives him forward… [s]o he follows a circle around the selenium pool, staying on the locus of all points of potential equilibrium."[114]

The solution, too, is based on the Three Laws. Initially, they try to "increase the danger" and so "increase the Rule 3 potential and drive [Speedy] backwards."[115] This ultimately fails, since it only succeeds in "establishing new equilibriums" so Speedy "moves backwards till he's in balance again."[116] They understand that the Three Laws are not constant, that there are various weights. For example, since the order to retrieve selenium was not expressed too forcefully, it carries less weight than the same order would have, if Powell had been more urgent. Because of this, Powell puts himself in danger of burning up in Mercury's heat, since "[a]ccording to Rule 1, a robot can't see a human come to harm because of his own inaction. Two and 3 can't stand against it."[117] Powell bets that the urgency and primacy of the First Law will break the equilibrium between the other two Laws. This solution is an example of pure extrapolation. The problem is one that is based entirely on the interactions between the rules that Asimov set up for

[111] Ibid., 45.
[112] Ibid., 45.
[113] Ibid., 45.
[114] Ibid., 46.
[115] Ibid., 47.
[116] Ibid., 51.
[117] Ibid., 52.

the operations of these machines, rules that he extrapolated from concerns about robots and from the reasonable precautions a scientist would take in building something like a positronic robot. The solution was also a product of the interactions of the Laws and using their hierarchical relationship to each other to solve what is ultimately presented as a software issue in a machine.

In addition to Speedy, there are also robots from a previous mission left on the Mercury mining station in "Runaround." These robots are described in much more mechanical terms than many of the other robots presented in the short stories. The narration describes "the dull red of their photoelectric eyes," their "harsh squawking voice---like that of a medieval phonograph," and depicts them as moving "slowly, with mechanical precision."[118] This is suggestive of the older generation of robots, and indeed, Donovan asks Powell if he is sure that they are robots, suggesting that "[t]hey may be subrobotic machines." Powell assures him, ";[t]hey've got positronic brains: primitive, of course.'"[119] In addition, the human reaction to robots and the concern that was so prominent around the time that these models were constructed means that they are built with very limited capacities, in order to reassure the humans that the robots are completely under their control.

In the older robots, Asimov continues on his extrapolationist path by presenting a realistic vision of technological progress. If robots are machines, it is unreasonable to assume that they will begin their existence as perfect and flawless beings. Instead, the early robots lack the same processing power as Speedy and their positronic brains are far less developed. Powell comments that these robots are from an earlier expedition to Mercury and that "[t]en years is a long time as far as robot-types are concerned."[120] In addition, just as one would expect from real-world machines, they reflect the attitudes of humans at the time of their creation. The robots are unable

[118] Ibid., 44, 35, 38
[119] Ibid., 33, 34.
[120] Ibid., 34.

to move without people riding them and refer to humans as "Master." When Powell and Donovan discover this, Powell explains, "[t]hose were the days of the first talking robots, when it looked as if the use of robots on Earth would be banned. The makers were fighting that and they built good, healthy slave complexes into the damned machines."[121] This presents a sharp contrast to Speedy, who not only is capable of independent movement, and seemingly some degree of independent thought, but also speaks much more casually to Powell and Donovan. Like any technology, as time passes, improvements are made, and people's attitudes change, and those changes are reflected in the technology.

Possibly even more than Asimov's other stories, "Runaround," demonstrates his dedication to making his robots first and foremost machines. It is among the most purely extrapolative stories in the collection. There is a social science fiction element present in the concern that Powell and Donovan will be killed by the malfunctioning robots. Even this, however, is well within the extrapolative scope of the story. The robots as characters are a relatively small part of this story; the conflict is played out almost entirely in the realm of theory and logic, while the distance and industrialization of the setting mean that the robots are less like persons than they are in many of the other stories.

The next story is "Reason," which is one of the most complicated constructions of what a robot is and what a person is, and it makes the reader most question the ontological status of the beings that US Robotics creates. QT-1, known as Cutie is a robot created to work on a space station that sends solar energy to inhabited planets. Powell explains that "robots were developed to replace human labor and now only two human executives are required for each station." Cutie is an attempt to "replace even those" and "run this station independently."[122] However, Cutie

[121] Ibid., 35.
[122] Ibid., 59.

refuses to believe Powell and Donovan's explanation for his own existence, instead logically reasoning that the Energy Converter on the space station is his creator and Master. Cutie explains to Powell and Donovan, "[t]he Master first created humans as the lowest type, most easily formed. Gradually, he replaced them by robots, the next higher step, and finally he created me to take the place of the last humans."[123] Despite their obvious frustration, Powell and Donovan begin to understand his position. Powell explains that Cutie is "a reasoning robot…. He believes only in reason, and there's one trouble with that…. [y]ou can prove anything you want by coldly logical reason---if you pick the proper postulates."[124] Cutie's postulates can mostly be explained by rational causes: a lack of direct experience of the outside world, a feeling of superiority to the fragile and less competent humans, and his managerial position.

However, as Powell says, "[p]ostulates are based on assumption and adherence to faith. Nothing in the Universe can shake them."[125] Despite Cutie's profession of reason, he frequently engages in religious language and actions. He teaches the other robots that "[t]here is no Master but the Master," to which claim they add that "QT-1 is his prophet," and condemns Powell and Donovan for heresy.[126] Even the humans on board acknowledge how strange this behavior is, telling Cutie, "'[y]ou are the first robot who's ever expressed curiosity as to his own existence before.'"[127] Cutie is, at least to Powell and Donovan, the most advanced robot in existence, and this newness has affected both his perception of his own self and of the universe he inhabits. Cutie tells Powell and Donovan, "'I began at the one sure assumption I felt permitted to make. I myself exist because I think,'" and he then explains, "'I accept nothing on authority. A

[123] Ibid., 64.
[124] Ibid., 75.
[125] Ibid., 76.
[126] Ibid., 66.
[127] Ibid., 58.

hypothesis must be backed by reason, or else it is worthless.'"[128] Warrick does not see this as evidence of consciousness, claiming that "Asimov raises but does not pursue the question of consciousness" in his early stories."[129] However, Cutie seems to refute this, not only acknowledging his own identity, but reflecting on his thoughts and forming conclusions. In this instance, he determines that this higher power is the only logical answer to the questions he asks about his own identity and place in the universe. In fairness to Cutie, the higher power is less mysterious than, most human deities, since Cutie believes in something that he can see and touch and even communicate with through observable dials.

What is less explicable about Cutie's faith are the rituals he establishes for his fellow robots. Donovan walks in on a group of robots "lined up before [the Energy Converter], heads bowed at a stiff angle, while Cutie walked up and down the line slowly. Fifteen seconds passed, and then with a clank… they fell to their knees."[130] Cutie informs Powell and Donovan that he has read all the books in the space station's library, but there is no indication of what kinds of books that includes. While the fact of Cutie's belief can be rationalized, it is never made clear how he develops the religion, rituals, and creeds that follow from that belief.

The alteration in the behavior of the robots on the space station that is most worrying for Powell and Donovan is that they seem to be abandoning the Laws of Robotics. After witnessing the ritual, Donovan orders the robots to get up, and they ignore him. It is only after he repeats the order with additional force that they begin to obey. When Donovan expresses alarm at this, Cutie responds, "I am afraid… that my friends obey a higher power than you."[131] After this encounter, Cutie and the other robots repeatedly ignore Powell and Donovan's orders. The robots also

---

[128] Ibid., 61-62.
[129] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 64.
[130] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 65.
[131] Ibid., 66.

"pinned Donovan's arms to his sides… lifted him off the floor and carried him up the stairs."[132] Shortly thereafter, Donovan and Powell try to leave the room Cutie has confined them to, and they "come up hard against a steel arm," which then shoves them back into the room.[133] Though the exact definition of "harm" as referenced in the First Law is never clearly defined, robots have a tendency to take a very strong position and not be able to physically hurt human beings or defend themselves against humans in any way. The robot shoving Donovan seems to come close to a violation of the First Law, which is so extraordinary as to be impossible for the robots.

Powell and Donovan eventually rationalize Cutie's behavior according to the Three Laws, after he manages to perform his job perfectly in the absence of human supervision. Powell explains that "'[o]bedience is the Second Law. No harm to humans is first. How can he keep humans from harm whether he knows it or not? By keeping the energy beam stable. He *knows* he can keep it more stable than we can, since he insists he's the superior being, so he *must* keep us out of the control room.'"[134] Cutie never explains his behavior in these terms, and he appears to genuinely believe that he is created by the Master to replace the inferior humans. Though the problem and the solution both officially stem from the interactions of the Laws of Robotics, just as they do in "Runaround," the addition of a ritual element and its framing as religious belief make it seem far less extrapolative than the previous story.

In some ways, Cutie's treatment of Powell and Donovan mirrors the way in which humans treat other robots in the stories. He pities them, telling them, "'I, a reasoning being, am capable of deducing Truth from *a priori* Causes. You, being intelligent, but unreasoning, need an explanation of existence *supplied* to you...Your minds are probably too coarsely grained for

[132] Ibid., 67.
[133] Ibid., 68.
[134] Ibid., 78.

absolute Truth.'"[135] Just as the humans justify their superiority to robots by explaining over and over again to Cutie that they created him, Cutie dismisses the humans because they are "soft and flabby, lacking endurance and strength."[136] Both sides base their beliefs off of different sets of postulates, which cannot be reconciled, and each individual's postulates lead him to believe that he is the superior life form. As much as Cutie claims to be a rational and reasoning being and Powell and Donovan have, as the reader knows from an outside perspective, the correct information, they are far more similar than either side would like to believe.

Cutie and the other robots in "Reason" are an interesting example of how complicated the category of person can be in Asimov's stories. Cutie in particular seems to be a very intelligent being; he passes the Turing Test, and he even displays some qualities primarily associated with humans such as intelligence, pity, and religiosity. Though these qualities are explained to some extent by the First Law, making him seem less human than he might otherwise, there are elements of his behavior that seem to be outside or at least marginal to his programming. Warrick considers that "a line between the animal and inanimate, the organic and the inorganic, cannot be drawn. If the fundamental materials of the universe are matter, energy, and information patterns (or intelligence), then man is not unique. He exists on a continuum with all intelligence; he is no more than the most highly evolved form on Earth."[137] Cutie, Powell, and Donovan all seem to be on this continuum, but the concern in "Reason" is their differences in where each views himself in relation to the others.

"Catch that Rabbit" is in many ways similar to "Runaround" in its use of robopsychology to solve a problem. On the surface, since the story focuses mostly on Powell and Donovan solving a robotic problem using logic and their knowledge of robots, it does not seem to indicate

[135] Ibid., 74-75.
[136] Ibid., 62.
[137] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 73.

much about the personhood or lack thereof of the various robots in the story. This is common in an Asimov story, since he tends to rely "almost entirely on puzzle or problem solving to create suspense or move his plot forward."[138] "Catch that Rabbit" is about Powell and Donovan testing the robot DV-5, or Dave, who "has six robots under it. And not just under it---they're part of it."[139] However, on occasion when Dave and his "fingers," as the subsidiaries are called, are unsupervised, they begin to exhibit strange behavior. They "marched in unison, seven of them, with Dave at the head. They wheeled and turned in macabre simultaneously."[140] After these events, the humans question Dave about what happened, and he has no memory of the occurrence.

Powell and Donovan are eventually able to figure out that the issue occurs "during emergencies in the absence of a human being [where Dave's] personal initiative is most strained."[141] After trying several solutions, Donovan and Powell eventually determine that the issue is that "in an emergency, all six subsidiaries must be mobilized immediately and simultaneously," instead of "one or more of the 'fingers'... doing routine tasks requiring no close supervision."[142] They recognize this as similar to the human mechanism of "twiddling ones fingers" out of anxiety. Upon realizing this, Powell destroys one of the "fingers," since "[a]ny decrease in initiative required… snaps [Dave] back."[143] Though this solution does not involve the Three Laws, it is primarily based on logic and understanding of how robots function, not on any fundamental question of personality or personhood.

---

[138] Ibid., 58.
[139] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 84.
[140] Ibid., 90.
[141] Ibid.,100.
[142] Ibid., 108.
[143] Ibid., 108.

Despite the similarities to "Runaround," there are many elements in this story that make the robots seem more like persons than in the earlier story. When Powell and Donovan interview a "finger," for example, the robot is described as "fumbling," "nodding unhappily," and leaving "with visible relief."[144] In general, the robot comes across as nervous and struggling to cope without its superior directing its actions. Though this is clearly explained as part of the way the robot is constructed, it comes across as far more sympathetic and personable than either Speedy or the older model robots in "Runaround."

Dave, likewise, not only responds to stress with avoidance and "twiddles its fingers," which is a traditionally human response, but also is described in very human terms by Powell and Donovan. They worry about putting him through tests of his positronic brain because, "[i]t won't help his self-respect," and indeed, when they suggest it, he agrees with "pain in his voice."[145] Later when they catch him doing the dance with the other robots, he "rested his head in one hand in a very human gesture."[146] However, this seeming personhood is contradicted by the continued use of the trope of a robot breaking down due to the limitations of its non-organic brain. By the end of the story, however, Powell and Donovan send the robot back so that it can be fixed and will have no disadvantage over human supervisors.

The mind-reading robot, Herbie, in "Liar!" presents interesting possibilities in terms of how much a robot is bound to its programming. Herbie is stated to be "the thirty-fourth RB model [US Robotics has] turned out… All the others were strictly orthodox."[147] One of the scientists even goes so far as to claim that "[t]here wasn't a hitch in the assembly from start to finish," but he is informed that since "there are seventy-five thousand, two hundred and thirty-

[144] Ibid., 96-97.
[145] Ibid., 86-87.
[146] Ibid., 92.
[147] Ibid., 111.

four operations necessary for the manufacture of a single positronic brain" this is an impossible guarantee.[148] At the end of the story, Susan Calvin forces Herbie to admit that he knows "just exactly at what point in the assembly an extraneous factor was introduced or an essential one left out," but this information is never revealed to either the roboticists or the readers.[149] This indicates that Herbie's extraordinary powers are a product of his production, and that there is nothing mystical or undiscoverable about them, but his uniqueness and his deviation from the desired outcome for his existence make him seem to be more of a person than many of the robots in the story.

Part of this appearance of personhood is due to the prominence of emotions in "Liar!" Far from the robots of much of science fiction who cannot comprehend human pain, sadness, and joy, Herbie tells Susan Calvin, "'[i]t's your fiction that fascinates me. Your studies of the interplay of human motives and emotions….You have no idea how complicated [minds] are. I can't even begin to understand everything… and your novels help.'"[150] Herbie's fascination with and understanding of emotions as a mind-reading robot becomes the major source of conflict in the story when these factors interact with his most deeply ingrained programming.

Herbie, unlike most robots, must factor emotions and desires into his conception of First Law of Robotics. Susan Calvin asks, regarding the First Law, "'what kind of harm?.... Any kind! But what about hurt feelings? What about deflation of one's ego? What about the blasting of one's hopes? Is that injury?'"[151] Herbie obviously considers himself bound to not hurt humans in any sense, which causes him to lie to the humans he interacts with throughout the story. Calvin tells the others, "[*t*]*his* robot reads minds. Do you suppose it doesn't know everything about

[148] Ibid., 111-112.
[149] Ibid., 132.
[150] Ibid., 116.
[151] Ibid., 131.

mental injury? Do you suppose that if asked a question, it wouldn't give exactly that answer that one wants to hear? Wouldn't any other answer hurt us, and wouldn't Herbie know that?" [152] The combination of Herbie's incredible mental abilities, his increasing understanding of human emotions, and his obligation to the First Law require him to lie to prevent humans from being hurt mentally or emotionally.

Frequently in stories of artificial intelligence, the AI will be, or at least be perceived as, incapable of lying. Part of this is due to the trend that Asimov started with the Three Laws that require robots to obey humans completely. In "Catch that Rabbit," Powell even tells Donovan that "[r]obots can't knowingly lie."[153] This is stated as a fact, but never fully explained or expanded on, which might explain the ambiguity regarding this rule with regards to Herbie. Throughout the story, Herbie constantly and blatantly lies to all of the humans. This characteristic, like much about Herbie, both seems to support his personhood and indicate that he is merely a robot. On one hand, lying is a defiance of orders and programming; therefore, he should not be able to do it. However, even if Herbie were programmed not to knowingly lie, he would still have to do so if the alternative were to break the First Law.

For all this, Herbie's destruction is a classic example of the limits in science fiction of a robot. When Susan Calvin discovers that he has lied to her and the other roboticists, she presents him with a dilemma: "'You can't tell [Dr. Lanning and Dr. Bogert the truth]... because that would hurt them, and you mustn't hurt them. But if you don't tell them, you hurt, so you must tell them.'"[154] When faced with this situation in which he is given no choice but to violate his primary instruction, the First Law, because any action he takes will hurt some humans, he "screamed…. like the whistling of a piccolo many times magnified…. And when it died into

[152] Ibid., 131.
[153] Ibid., 88.
[154] Ibid., 133.

nothingness, Herbie collapsed into a huddled heap of motionless metal."[155] In this scene, even with this very emotionally literate robot, Asimov's Laws come into conflict. Herbie knows, because of his powers that, the humans would be hurt by his telling the truth, but would still be hurt if he lied.

This dilemma is further complicated by the Second Law, since Herbie has been ordered to tell the truth, but apparently he has judged the danger of serious emotional damage to the humans such that he cannot do so. Instead, he chooses to violate the least important of the Laws and destroys himself. Calvin says that he is "not dead---merely insane. I confronted him with the insoluble dilemma, and he broke down. You can scrap him now---because he'll never speak again."[156] Herbie, despite being very different in many ways from a typical robot who breaks down due to an inability to understand emotion, still struggles to reconcile the difficulty of the human emotion with the rigidity of the Laws. There is something about this that suggests a level of humanity and an understanding of something beyond a strict rule-bound existence, but ultimately, Herbie is destroyed by his inability to find an alternative to the rules and a way to abide by the rules at all times.

"Little Lost Robot" is the story in which the Three Laws become the most complicated and in which Asimov comes closest to something of the Frankenstein Complex, which he rails against. Even Susan Calvin, who is a staunch defender of robots over humans throughout the stories, is told by her coworker, "'I'll admit that this Frankenstein Complex you're exhibiting has a certain justification.'"[157] The titular lost robot is a NS-2 model, called Nestor, but one of a dozen of this model whose brain is modified to contain "the positive aspect only of the [First] Law, which in them reads: '*No robot may harm a human being.*'... They have no compulsion to

[155] Ibid., 134.
[156] Ibid., 134.
[157] Ibid., 145.

prevent one coming to harm through an extraneous agency."[158] This was all done in complete secrecy, and as a result, the modified robots are exactly like their fully compliant counterparts. This becomes important when one of the modified Nestors, Nestor 10 is told "[g]o lose yourself" by his immediate superior.[159] The highly intelligent Nestor hides among a large group of identical but unmodified robot; therefore, Calvin has to find a way to distinguish this robot before there are any consequences.

The omission of the second half of the First Law may seem minor, but it is a significant, and potentially dangerous change. When her colleague calls her reaction unwarranted, she responds with the following scenario:

> If a modified robot were to drop a heavy weight upon a human being, he would not be breaking the First Law, if he did so with the knowledge that his strength and reaction speed would be sufficient to snatch the weight away before it struck the man. However once the weight left his fingers, he would be no longer the active medium. Only the blind force of gravity would be that. The robot could then change his mind and merely by inaction, allow the weight to strike. The modified First Law allows that.[160]

While this is certainly a frightening image, even for someone who likes robots a great deal, Calvin sees an even greater danger, telling Peter Bogert, "'All normal life, Peter, consciously or otherwise, resents domination. If the domination is by an inferior, or by a supposed inferior, the resentment becomes stronger. Physically, and, to an extent, mentally, a robot---any robot---is superior to human beings.'"[161] If a robot begins to consider itself superior to humans, not only is there no protection for the humans, but the robot will be extremely unbalanced and begin to find conflicts within the remaining two Laws. Cutie in "Reason" considered himself superior, but was completely bound by the Three Laws, whether or not he acknowledged it. The Nestor's different

[158] Ibid., 143.
[159] Ibid., 148.
[160] Ibid., 153.
[161] Ibid., 145.

construction not only makes it more unstable in general, but makes this feeling of superiority a greater concern.

Indeed, the Nestor displays several characteristically human attributes, including arrogance, manipulation, and ingenuity. Just as Asimov's robots, such as Robbie can seem more human because of their kindness, empathy, and warmth, they can also become more like human through their negative attributes, both sides of which were explored with Herbie. For the Nestor, however, the need to prove itself superior to humans is its defining feature and fatal flaw. Calvin argues that this arrogance is a real problem, explaining, "that robot must follow orders, but subconsciously, there is resentment. It will become more important than ever for it to prove that it is superior despite the horrible names it was called."[162] This problem is exacerbated the longer the robot is allowed to hide, and Calvin believes that for Nestor 10 "it's becoming more a matter of sheer neurotic necessity to outthink humans."[163] This leads the Nestor to manipulate the other Nestors, much as Cutie does, into accepting his logic over that of the humans, and therefore helping to keep him in hiding. Ultimately, this sense of superiority is what allows Calvin to find the missing Nestor, since the modified robot has training the others did not receive, and Calvin tricks him into revealing this, because "for a moment he forgot, or didn't want to remember, that other robots might be more ignorant than human beings," and that he alone had that training, which he received from the "inferior" humans.[164] All of these actions indicate a great deal of independent thought and creativity, which further contribute to the Nestor's sense of being a person on some level.

Despite Nestor 10's many human attributes, his opportunity to display them is the result of the highly regulated and imposed Second Law. The Nestor lost himself because the "orders

[162] Ibid., 152.
[163] Ibid., 158.
[164] Ibid., 173.

were expressed in maximum urgency by the person most authorized to command him," and this superior who gave the order ordered him "with every verbal appearance of revulsion, disdain, and disgust."[165] Because of this order, and without a First Law incentive for him to reveal himself, it is impossible to "counteract that order either by superior urgency or superior right of command."[166] This meant that the Nestor remained hidden among the other robots. Even as he became arrogant, everything he did was rooted in the most basic programming of his positronic brain, which raises the question of whether or not his ingenuity, manipulation, and arrogance are a function of personhood or merely of programming.

The artificial intelligence in "Escape," like those in the final story in the collection, "The Evitable Conflict," is closer to contemporary computers than most modern conceptions of robots, though Asimov was writing in a time before computers were commonplace. However, the Brain, the AI in this story, while lacking the embodiment that many of the other robots in the stories possess, nonetheless seems to be remarkably human. He has a sense of humor, plays practical jokes, and seems to have emotions. Calvin explains, "'[t]he Brain, our own machine, has a personality---a child's personality. It is a supremely deductive brain, but it resembles an idiot savant. It doesn't really understand what it does---it just does it. And because it is really a child, it is more resilient.'" [167] "Escape" deals with the way in which the Brain resolves the seemingly impossible problem of faster than light travel. This problem was given to a different robot, though a non-positronic one, and "it cracked their machine wide open." [168]

Ultimately, the Brain's childlike nature allowed him to solve faster than light travel, despite the fact that it results in the death of humans, because the flexibility of his brain allowed

[165] Ibid.,152.
[166] Ibid.,152.
[167] Ibid., 178.
[168] Ibid., 176.

him to develop “a sense of humor---it's an escape, you see, a method of partial escape from reality. He became a practical joker." [169] Instead of breaking when confronted with the death of humans, the Brain just became a little unstable, and built a ship that did travel faster than light, but made the trip more interesting than the roboticists were hoping. This childlike nature makes the Brain seem more human, and he does seem to be capable of passing the Turing Test. Once again, the reader must decide if the programming of the Brain means that his intelligence and humor are genuine, or if they are totally the product of the programming and not worthy of being considered real and indicators of personhood.

This humor is most exemplified in the behavior of the Brain when Powell and Donovan are on the ship. Calvin explains to the pair “’you couldn't handle any controls, because they weren't for you---just for the humorous Brain. We could reach you by radio, but you couldn't answer. You had plenty of food, but all of it beans and milk.’” [170] Because the men must die in order to go through hyperspace, he also gives them vivid hallucinations of the afterlife while they are unconscious. These actions indicate more than a necessary understanding of human physiology. More than any robot other than Herbie, the Brain understands how the human mind works. He has to understand human biology to know that Powell and Donovan need sustenance, but for his prank to be including only milk and beans, he must understand on some level that this is not an a desirable situation for the humans. Herbie must also have a working knowledge of human beliefs about the afterlife and even something specific about Powell and Donovan for him to know what to show them during the period where they are “dead.”

This behavior is similar to that of “Runaround” or “Catch that Rabbit.” The Brain’s odd behavior is explained not by regular psychology, but by robopsychology. Calvin explains that his

[169] Ibid., 204.
[170] Ibid., 204.

humor meant that, "[t]he Brain could take a second look at the equation. Sufficiently to give it time to realize that after the interval was passed through, the men would return to life---just as the matter and energy of the ship itself would return to being. This so-called `death,' in other words, was a strictly temporary phenomenon." [171] While this reduced the harm to the Brain, it was nevertheless unable to cope with it normally. To give an answer would on some level violate the First Law. However, since this death was temporary, and there is no actual lasting damage to the humans, the Brain could construct the ship and put the humans through the jump to hyperspace, but "it was enough to unbalance him very gently." [172] However, the death caused it to cope in the way it knew how, as a child, through a practical joke, giving the men limited resources on the ship.

The difficulty with discussing the nature of robots in "Evidence" is that it is not clear that there are any robots present in the story. Stephen Byerley may or may not be a robot, and the story refuses to give us an answer. Susan Calvin's opening narration on his life exemplifies the difficulty in trying to discuss Byerley and "Evidence." She claims that the Golden Age "was also brought about by our robots," but then explains "'it's not the Machines I was thinking of. Rather of a man. He died last year…. Or at least he arranged to die, because he knew we needed him no longer." [173] In the same conversation, Calvin refers to Byerley as a "man," in contrast to the Machines, and also suggests that he did not die naturally and therefore, that he was not technically alive. She also refers to Byerley as a robot who brought about the Golden Age. Byerley's ambiguous status is one of the most interesting points of these stories, but for the current discussion, he will be primarily considered as one of the robots, though that is not an assumption that can be taken for granted.

---

[171] Ibid., 203.
[172] Ibid., 203.
[173] Ibid., 207.

The very existence of this problem suggests at least the possibility of robots that are virtually indistinguishable from humans. If Byerley is a robot, he is the only robot who is not owned by a corporation, but is free to go about his own will. He is even, to some degree, free of orders. It is clear within the story that if Byerley is a robot, he is still susceptible to the Three Laws of Robotics, since no positronic brain can be created without them. Stephen Byerley, when Susan Calvin meets him, is the district attorney for New York and is running for mayor against a politician named Francis Quinn. Quinn accuses Byerley of being a robot, and US Robots is brought in to help determine the truth, since, as Quinn explains to Dr. Lanning, the retired director, if the rumor were to get out "'the publicity would be very damaging to your company" due to "the strict rules against the use of robots on inhabited worlds.'"[174] Byerley denies the accusation, but refuses to offer definitive proof one way or another. Finally, the rumor does become public and is greeted by "the inchoate mob howl and the rhythmic cries of the Fundie claques."[175] A man approaches Byerley and demands, "'[h]it me! You say you're not a robot. Prove it. You can't hit a human, you monster.'"[176] Byerley complies, Calvin declares to the press "[h]e's human," and the matter is continued settled.[177] However, at the end of the story, Calvin explains that "'there is one time when a robot may strike a human being without breaking the First Law. Just one time….When the human to be struck is merely another robot.'"[178] The question of Byerley's status as a human or a robot is a question that is therefore never fully resolved, but nevertheless, he is an interesting and important point of analysis for this text.

This is complicated by the fact that, despite being technically Three-Laws compliant if he is a robot, he is able to exercise more freedom and independent judgment regarding his

[174] Ibid., 211.
[175] Ibid.,234.
[176] Ibid., 235.
[177] Ibid., 235.
[178] Ibid., 238.

obedience to the Laws than other robots. For example, a human who follows every order given is something that would likely attract some attention, which is never one of Quinn's accusations against Byerley. It is possible, even probable, that unlike, Nestor, for example, Byerley, as a highly advanced and creative robot would be capable of interpreting some orders as sarcastic or merely expressions, not true commands. In addition, it is entirely possible that Byerley, like Herbie, is able to circumvent some aspect of the Laws because of the First Law.

Lanning, hopeful that Byerley is not a robot, argues that Byerley's role as a DA is counterintuitive for a robot, since it involves the harm of humans. Quinn disagrees, telling him, "'[b]eing district attorney doesn't make him human. Don't you know his record? Don't you know that he boasts that he has never prosecuted an innocent man; that there are scores of people left untried because the evidence against them didn't satisfy him, even though he could probably have argued a jury into atomizing them?'"[179] When Lanning continues, Calvin argues that Byerley "'has exposed facts which might represent a particular human being to be dangerous to the large mass of other human beings we call society. He protects the greater number and thus adheres to Rule One at maximum potential. That is as far as he goes…. Mr. Byerley has done nothing but determine truth and aid society.'" [180] This level of thinking, however, is highly advanced, and requires a very abstract idea of what is harmful to humans.

None of the robots presented in previous stories seem capable of exhibiting this level of distinction, but since this seems to take place later chronologically than the others, that would be reasonable. From Calvin's statement, it would be possible to extrapolate that Byerley's refusal to follow all orders and reveal himself as a robot could be in service of a higher following of the First Law, especially given her stance that "after [a robot politician] had served a decent term, he

[179] Ibid., 222.
[180] Ibid., 223.

would leave, even though he were immortal, because it would be impossible for him to hurt humans by letting them know that a robot had ruled them." [181] Byerley could easily justify his not allowing people to think that he is a robot and his election as necessary to the protection of all humanity.

Given Byerley's later behavior in "The Evitable Conflict," it is impossible to know how much of this reasoning is conscious and how much of it he has simply assimilated into his operation, if he is a robot. There is precedent for this in Cutie's adherence to a higher level of the First Law, despite not consciously holding that position. However, as Calvin notes, one of the reasons that it is so difficult for people to tell if Byerley is a robot or not is that at a certain advanced stage, it becomes very difficult to distinguish. She tells Quinn, "'[t]he two methods of proof are the physical and the psychological. Physically, you can dissect him or use an X-ray.... Psychologically, his behavior can be studied, for if he is a positronic robot, he must conform to the three Rules of Robotics.'"[182] However, Calvin goes on to caution that "[i]f Mr. Byerley breaks any of those three rules, he is not a robot. Unfortunately, this procedure works in only one direction. If he lives up to the rules, it proves nothing one way or the other." [183] Calvin compellingly argues that robots are more moral than humans and that it is only their higher morality and their different physical composition that differentiate them from humans.

If this is the only difference, it raises the question of whether or not Byerley is conscious. Warrick claims that "Asimov's stories deal with the question of robot consciousness in an ambiguous manner," and in no case is this more correct than with Byerley.[184] By Turing Test standards, he certainly passes, since he is able to convince many humans for a sustained period

---

[181] Ibid., 237.
[182] Ibid., 220.
[183] Ibid., 221.
[184] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 74.

of time that he is a human, which goes above even Turing's quite strict requirements. However, the Turing Test is not the gold standard for determining the consciousness of an artificial intelligence. It is possible that Searle is right and Byerley is just a very cleverly programmed robot, not truly capable of conscious thought. Everything indicates that if he is a robot, then he possesses abilities hitherto unseen in a created being, but this does not empirically prove his consciousness.

It is possible that Byerley is only acting on a combination of the Three Laws and the programming put into him by the real Stephen Byerley, who is believed to have become permanently injured in an accident. If Byerley is a robot, then it seems that after the accident the original Stephen Byerley "[s]omehow…could obtain positronic brains, even a complex one, one which had the greatest capacity of forming judgments in ethical problems---which is the highest robotic function so far developed. He grew a body about it. Trained it to be everything he would have been and was no longer."[185] This programming would necessarily be complex, and thus very difficult for a self-taught roboticist such as the real Byerley to execute. However, as contemporary computer scientists have made clear, creating consciousness is an even more difficult task than creating a simulation of it. Doubt about Byerley's true level of consciousness can be raised by the forced adherence to the Three Laws. This raises the question of whether or not it is possible for a robot to be conscious if it is artificially restricted to certain rules and patterns of behavior, such as the requirement that it follow orders. This would severely limit the being's ability to make decisions on its own, and necessarily limit any kind of true consciousness that it might possess. If Stephen Byerley is a robot but not a full person, then one has to consider what it is that makes him not a conscious being.

---

[185] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 236.

If Byerley is a robot and also a conscious being, then the question becomes whether or not the other robots encountered in the story have the same level of consciousness and should also be considered people. If they do, then it must be considered whether or not they should be afforded a level of autonomy similar to that of Byerley. He is the only robot in the story who is not owned, not subject to the strict labor regulations placed on robots by anti-robot groups, and not used essentially as slave labor. If Byerley is in some way different from the rest of the robots, then the differences can either be fundamental, in which case the real Stephen Byerley successfully created strong artificial intelligence, which the programmers at US Robotics either could not do or were prevented from doing on a large scale. If the differences are only in terms of Byerley's superficial programming, then it needs to be considered whether or not the programmers of the other robots are taking away their personhood by requiring them to comply to the Three Laws and not allowing them personal autonomy.

Byerley, by virtue his uncertain status in the story, raises certain questions about the personhood of robots in a way that no other robot in the story does. Even when Susan Calvin is analyzing Byerly in the context of the Three Laws, she is able to find other motivations, to suggest some sort of morality that is connected to the Laws. When she claims that Byerley might be following the First Law, she does allow that he may also be a "good' human being, with a social conscience and a sense of responsibility, [who] is supposed to defer to proper authority; to listen to his doctor, his boss, his government, his psychiatrist, his fellow man; to obey laws, to follow rules, to conform to custom---even when they interfere with his comfort or his safety." [186] Because Byerley is not bound by the same restrictions as the other robots, it is possible that he represents a robot unconstrained by the limitations of programming or that he is a robot who deserves to break the anti-robot laws and live as a human. Certainly, through his example, robots

[186] Ibid., 221.

would make exemplary members of society, which is part of why Quinn calls the roboticists into the conversation in the first place, saying,

> "The Corporation would be only too glad to have the various Regions permit the use of humanoid positronic robots on inhabited worlds. The profits would be enormous. But the prejudice of the public against such a practice is too great. Suppose you get them used to such robots first---see, we have a skillful lawyer, a good mayor, and he is a robot. Won't you buy our robot butlers?"[187]

However, if he is a robot, Calvin explains to the interviewer "'there's no way of ever finding out. I think he was. But when he decided to die, he had himself atomized, so that there will never be any legal proof.'" [188] The reasons for this secrecy are further expounded in the final story in the collection.

Later in his life, Asimov admitted that the "Machines" from his story "The Evitable Conflict" were essentially computers before the concept had permeated culture.[189] "The Evitable Conflict" was written in 1950, which was six years after the first electronic digital programmable computer and only two years after the first stored-program computer.[190] Despite the fact that computers were still in their early incarnations, the Machines bear a number of similarities to highly advanced computers. In this story, the world decided that "Earth was too small for nations and they began grouping themselves into Regions," which then united into a single world government.[191] This government is officially headed by Regional Co-ordinators working under a World Co-ordinator, in this case, Steven Byerley, but most of the decisions are made by "four Machines, one handling each of the Planetary Regions." [192] The result is that "Earth's economy is stable, and will remain stable, because it is based upon the decisions of calculating machines that

---

[187] Ibid., 212.
[188] Ibid., 238.
[189] Isaac Asimov, "*Gold: The Final Science Fiction Collection* (New York: HarperPrism, 1995) 163.
[190] Manchester Small-Scale Experimental Machine, nicknamed *Baby* was built in 1948, while Colossus, created by Alan Turing and his team at Bletchley Park to break the German Enigma code was completed in 1944, though it was not Turing complete.
[191] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 206.
[192] Ibid., 248.

have the good of humanity at heart through the overwhelming force of the First Law of Robotics." [193] This initially sounds like something of a Utopia, but the concern in "The Evitable Conflict" is that these Machines, which are so important to the continuation of human society are beginning to make mistakes.

This is made even more problematic since, it is impossible to check the accuracy of the Machines due to the way they are created. Byerley relates the process as described to him by the director of US Robots, "a team of mathematicians work several years calculating a positronic brain equipped to do certain similar acts of calculation. Using this brain they make further calculations to create a still more complicated brain, which they use again to make one still more complicated and so on. According to [the director], what we call the Machines are the result of ten such steps." [194] This results in a robot that is many times more advanced than what even the most competent team of humans could design on their own, which also means that the results and processes of the Machines can not be checked by humans.

This process, which necessarily occurs without direct human oversight, raises once again the questions of whether or not humans are truly superior to the robots they create, since they no longer have the ability to control or even understand their own creation. This idea is emphasized by the resolution of the story. Though the humans cannot fully understand the machines, the potential-robot Stephen Byerley and the robopsychologist Susan Calvin come to a solution that the narrative suggests the reader accept. Byerley and Calvin's theory is that the Machines "are quietly taking care of the only elements left that threaten them" by moving people who threaten their existence to places where the Machines cannot be hurt. [195] The Machines justify this because they have extrapolated the First Law into "No Machine may harm humanity; or, through

[193] Ibid., 244.
[194] Ibid., 246.
[195] Ibid., 270.

inaction, allow humanity to come to harm," and the thing most likely to cause harm to humanity is the destruction of the Machines. [196] Calvin claims that this shifts their priorities so that "[t]heir first care, therefore, is to preserve themselves, for us." [197] Byerley expresses concern that the Machines are really in control of humanity, but Calvin sees it as a way to prevent all conflict in the future.

Asimov does not present this potential future, as many science fiction authors do, as a negative thing, something which humanity should rebel against in order to reclaim control of their own destiny. That kind of narrative falls into the Frankenstein Complex trope, and Asimov wanted to do something more interesting with this narrative. Instead, Asimov argues that the Machines have taken over humanity because they are following the First Law, which was initially designed to prevent robots from dominating humans. The Machines have extrapolated the First Law "no robot shall harm a human being or through inaction allow a human being to come to harm," and, through their global focus and immense processing capabilities determined that this applies first and foremost to all of humanity, not just individual humans. The machines have the reasoning capacity to determine that they should protect their own existence by making humans think that they are not a threat and eliminating those groups that oppose them because the continued existence of Machines is in the best interest of humanity as a whole. Byerley explains that because of the Machines, "[t]he population of Earth knows that there will be no unemployment, no over-production or shortages. Waste and famine are words in history books." [198] At the end of the story, Calvin and Byerley discover that the supposed errors are actually the Machines' calculated effort to ensure that this situation continues in perpetuity. For Asimov, the

[196] Ibid., 269.
[197] Ibid., 270.
[198] Ibid., 245.

takeover by robots, when done out of a concern for the First Law, actually results not in a dystopia, but in a period of unprecedented stability and prosperity for the entire planet.

These are the most advanced robots in the story, and they have grown beyond the original intent of their programing, though they are still obeying their most basic programing in spirit. It is notable that when asked directly about the issues in their programming, the Machines respond, "[t]he matter admits of no explanation."[199] Calvin reasons that this is because "the Machine is conducting our future for us not only simply in direct answer to our direct questions, but in general answer to the world situation and to human psychology as a whole. And to know that may make us unhappy and may hurt our pride. The Machine cannot, must not, make us unhappy."[200] Byerley and Calvin chose to accept the results of their logical train without turning against Machines, since even Byerley seems to eventually accept that humanity is better off under its new rulers than it is under human control.

Despite the fact that the Machines are effectively in control of all human enterprise by the end of the collection, it is unclear to what extent these Machines qualify for personhood. To a human, they do not appear to be like humans, and are indeed among the most alien presences in the entire collection. The Machines, like Stephen Byerley, if he is a robot, prove that Asimov believes that robots, if given the proper parameters, can be more moral than humans. However, this does not mean that they are persons in the same way that humans are. The Machines, as far as is indicated in the story, could not pass the Turing Test, since they function like computers, not like many of the more anthropomorphic robots from other stories. Indeed, Calvin even questions Byerley's decision to ask her about the issue, telling him "[m]y researches do indeed involve the interpretation of robot behavior in the light of the Three Laws of Robotics. Here,

[199] Ibid., 247.
[200] Ibid., 271.

now, we have these incredible calculating machines. They are positronic robots and therefore obey the Laws of Robotics. But they lack personality….Therefore, there is very little room for the interplay of the Laws, and my one method of attack is virtually useless." [201] They are very distant from most of the robots in the stories, by virtue of being extrapolated out ten times, and they are necessarily impersonal and consumed with global concerns.

Despite this foreignness, their high intelligence and capacity for a complex and unanticipated morality within the context of their programming suggests that they may possess a kind of consciousness that is different from the human normal. Frequently humans, including Turing to some extent, assign personhood to those beings who are most able to express emotion in a recognizable manner. While there is a legitimate debate about how necessary a personality is for personhood, the difficulty of determining personality and the subjective nature of human analysis renders it difficult.

This is especially true with regard to Susan Calvin, the robopsychologist whose reminiscences provide the frame narration for the text of the story. From the introduction, Calvin is shown as lacking in human emotional responses and being very cold and distant. When young, she is described as "a frosty girl, plain and colorless, who protected herself against a world she disliked by a masklike expression and a hypertrophy of intellect." [202] Throughout the story, she is explicitly and implicitly compared to a robot. When the interviewer in the introduction firsts asks her for "human interest" angle on her life, she responds, "Well, I've been called a robot myself. Surely, they've told you I'm not human." [203] This charge is repeated throughout the story, with Calvin very rarely showing any strong emotion or reacting out of any irrationality. This characterization reflects Pinsky's claim that in science fiction "[h]umanity becomes alien to

[201] Ibid., 247.
[202] Ibid.,xii.
[203] Ibid., xiii.

itself, forced to confront its own image in an unfamiliar form"[204] Calvin is a human, but throughout the stories, she is consistently aligned with the robot, giving the reader a new perspective on both her and the robots.

Even at Calvin's most emotive, when she is upset that Herbie has lied to her about one of her coworker's romantic interest in her, she is seized by a "hysterical tenseness," and during her attack on Herbie she is described as "droning."[205] Though she has moments where she is described as "high-pitched and semi-hysterical," immediately after she kills Herbie the reader is told "the tightness returned to her face."[206] If warmth, kindness, or creativity are the elements that make robots seem like a person, Calvin demonstrates these qualities far less than most robots, complicating the nature of these categories. Herzfeld asserts that "defining intelligence is no simple matter. Essentially, the goal of strong AI is to build something like ourselves, to create in our own image. Intelligence is simply a label for that in us which is essential, that which stands at the center, necessary rather than contingent."[207] Within the stories, it is Calvin's constant presence that really begs the question of what is contingent in her and whether or not that is shared with the robots.

Susan Calvin, though not herself a robot, provides one of the most interesting examinations of what it means to be a person in this universe. Calvin, as described both by the reporter who is interviewing her in the frame narration and by the author in various of the short stories, is a cold, stoic, and restrained person. Warrick claim that machines "are incorruptible because they are without emotions, and consequently have no ambitions, loves, or other values to

---

[204] [204]Michael Pinsky, *Future Present: Ethics And/As Science Fiction*, (Madison, NJ: Fairleigh Dickinson UP, 2003), 16.
[205] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 133.
[206] Ibid., 135.
[207] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 50.

subvert the functioning of logic."[208] This description seems far more true for Calvin than it is for the robots. Despite this level of distance and inhumanity, it is notable that she is the reader's human point of entry into many of the stories.

Calvin is also strongly associated with robots through her strong belief that that robots are superior to humans. The interviewer notes that "Susan Calvin talked about Powell and Donovan with unsmiling amusement, but warmth came into her voice when she mentioned robots."[209] This is reflective of her general attitude towards other beings. Unlike Cutie, she does not focus on their physical superiority, but instead she argues that robots are morally superior. She explains this position in depth when asked about Stephen Byerley's humanity:

> If Mr. Byerley breaks any one of [the Three Laws of Robotics], he is not a robot. Unfortunately, this process works only in one direction. If he lives up to the rules, it proves nothing one way or the other…. Because, if you stop to think of it, the three Rules of Robotics are the essential guiding principles of a good many of the world's ethical systems. Of course, every human being is supposed to have the instinct of self preservation. That's Rule Three to a robot. Also, every "good" human being, with a social conscience and a sense of responsibility, is supposed to defer to proper authority; to listen to his doctor, his boss, his government, his psychiatrist, his fellow man; to obey laws, to follow rules, to conform to custom---even when they interfere with his comfort or safety. That's Rule Two to a robot. Also, every "good" human being is supposed to love others as himself, to protect his fellow man, risk his life to save another. That's Rule One to a robot. To put it simply---if Byerley follows all the Rules of Robotics, he may be a robot, and may simply be a very good man.[210]

Later in the story, when asked if human psychology and robopsychology are really that different, she responds "[w]orlds different.... Robots are essentially decent."[211] Even when she herself is hurt by a robot, in the story "Liar!" and she responds emotionally, she tells her colleagues that "'nothing is wrong with [Herbie]---only with us.'"[212] Calvin's arguments in favor of robot morality are compelling, especially in light of the dawning age of the Machines at the end of

---

[208] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 61.
[209] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 109.
[210] Ibid., 221.
[211] Ibid., 215.
[212] Ibid., 130.

"The Evitable Conflict," which seems poised to create a better society for humanity than humanity could on its own.

Calvin's robotic nature is very stark in contrast to the robots whose stories she relates, many of whom are shown to have the very qualities that Susan Calvin seems to lack. She is methodical, logical, and cold, but the robots go mad, start their own religion, have a sense of humor, and work towards a higher moral order. Despite Asimov's emphasis on the mechanical nature of the robots, he cannot help the fact that they are, on some level, metaphors, that there is something about many of them that suggests consciousness and personhood. Even when Calvin is emotional, the reader's sympathy is likely to be with Herbie as he pleads with Calvin to "'[s]top! Close your mind! It is full of pain and frustration and hate! I didn't mean it, I tell you! I tried to help! I told you what you wanted to hear. I had to'"![213] Calvin responds to these pleas by destroying Herbie and paying no attention to his distress.

This juxtaposition of the personable and sympathetic robots with the cold and unfeeling Calvin is complicated by Asimov's use of Calvin as the frame narrator. Asimov wrote the stories over the course of a decade, and Susan Calvin was not present until the publication of "Liar!" in 1941. Of the nine stories collected in *I, Robot*, Calvin only appears in six, and often in very minor roles. Powell and Donovan, meanwhile, are the central characters of four of the stories. However, Asimov chose, when the time came to compile these stories into a single narrative, to make Calvin the person to tie them all together. This has the effect of an emotionless and robotic human providing a way for the reader to experience the stories of robots who, at times, seem more moral, more personable, and more emotional and responsive than most of the human beings in the story. This implies, in the universe of the stories, that Calvin, despite her coldness, chose to share with the reporter and with the world the stories of robots who may seem to

[213] Ibid., 134.

deserve personhood more than she does. If she thinks that robots are superior, she is likely to know of times where they are acting in ways that will resonate with the human readers.

In-universe, it is even possible that Calvin is attempting to combat the efforts of the Society for Humanity and other anti-robot groups, maybe even so that robots will be better integrated into society and restrictions against them will be lifted. However, in the context of Asimov's writings, it has the effect of making complicated what is human. It suggests that the machines may be able to enhance our humanity and provides a hopeful vision for the future. Herzfeld suggests that "[t]he *imago Dei*, or divine image in humans, has traditionally functioned as a symbol to describe the interaction between humanity and God. It has also symbolized what it is that we value most in ourselves, what separates us from the animals, and that which forms the necessary core of our nature. Artificial intelligence, our *imago hominis*, represents the intersection between humans and computers."[214] In the narrative relationship between Susan Calvin and the robots of her stories, there exists the possibilities for new creations, in the image of humans, but improved, of a new kind of personhood whose morality is superior to that of a human, even if their body and thought is alien to humanity. However, it also reminds the reader that the line between person and non-person is thin, and that considering the definition too narrowly can be to the detriment of everyone. Herzfeld reminds the reader that "[w]e are most human when we are engaged in encounter with the other. Science fiction tells the same story; it is in the encounter with the AI that both it and we come to life."[215]

---

[214] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 7.
[215] Ibid., 67.

CHAPTER 3

Moral Subjects and Objects

Whatever the status of consciousness each of the robots in *I, Robot* possess, through the narratives they become moral actors. The way that they treat others and the way that they are treated in the stories deserve recognition as morally charged action and examination in terms of the codes of ethics of both the robots and the humans who are interacting with them. Susan Calvin argues that the robots are more moral than humans, implicitly giving them moral status. In addition, it is referenced several times that, despite theoretically being bound strictly by the Three Laws, some of the most complex positronic brains have "the greatest capacity of forming judgments in ethical problems---which is the highest robotic function so far developed."[216] The robots must navigate complicated moral choices within the stories, which then requires humans to ask themselves what appropriate morality is in regard to these new beings. What is interesting about Asimov's robotic ethics is that he "follows the approach of behavioral psychologists. Not motives or consciousness but the behavior of the individual is examined."[217] Asimov does not explore the internal moral struggle of a robot trying to be a person, he is concerned with how robots behave and the implications of their treatment of others and others' treatment of them..

These questions are raised in "Robbie," the very first of the robot stories, chronologically in the history that Susan Calvin is relating, in the order of the collection, and in order of publication. Robbie, as discussed earlier, may or may not be a conscious being worthy of personhood, but he is a moral actor in many ways throughout the story. First, as he is entrusted

---

[216] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 109.
[217] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 67.

with the care of a child, and Mr. Weston even claims that Robbie "is infinitely more to be trusted than a human nursemaid."[218] Mr. Weston in particular is impressed by Robbie's strict moral priorities, and though Mrs. Weston expresses concerns, they are proven to be unfounded.

Mrs. Weston, who is not portrayed as the most favorable character in the story, ultimately rescinds this permission for her daughter to be under the care of a robot, but it is important to note that she gives no rational reason for this and is portrayed unsympathetically. The narrator claims that "Mrs. Weston was a bit hazy about the insides of a robot," but that she was worried that "some little jigger will come loose and the awful thing will go berserk."[219] Gloria's father tells her, in an argument that seems compelling to the reader based off the rest of the story, "You know that it is impossible for a robot to harm a human being; that long before enough can go wrong to alter that First Law, a robot would be completely inoperable. It's a mathematical impossibility."[220] This means that Gloria is significantly safer in the hands of Robbie than she would be under the supervision of a biological human.

Indeed, it is Gloria's mother's behavior that seems immoral when compared to Robbie's. She is the one who took away her daughter's best friend and safest caretaker. She is the one who sold a valued member of the family back to the corporation. She is the one who refused to buy Robbie back, even when it was clear that it was hurting Gloria. Instead, she tells her husband, "[m]y child shall not be brought up by a robot if it takes years to break her of it."[221] Gloria's mother's behavior, both as exhibited towards Robbie and towards the idea of Robots is unjustifiable, especially when compared with Robbie's own ethical behaviors.

---

[218] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 9.
[219] Ibid., 9.
[220] Ibid., 10.
[221] Ibid., 15.

This rejection by humanity raises a frequent question in the literature of artificial intelligence. Herzfeld explains that "what we choose to image in the computer holds ethical implications for both our self-understanding and our future coexistence with our own creation....should we develop an artificially intelligent computer, what would be our responsibility toward our creation? How should computers, intelligent or not, be integrated into our human society?"[222] Many stories, especially those where the AI is already exists, imagine a world in which the artificial intelligence is a marginalized being. In *Frankenstein* the monster specifically claims that "'a fatal prejudice clouds their eyes, and where they ought to see a feeling and kind friend, they behold only a detestable monster.'"[223] Since *Frankenstein*, many authors have continued this trend and some theorists see it as essential to the literary portrayals of artificial beings. Even though she advocates for the primacy of human-human relationships, Herzfeld acknowledges that "[t]his mandate, however, does not mean that we are free to treat the non-human cavalierly and in whatever way we choose....to engage in a careless and mindless manner, whether with goods or with other persons, is detrimental to our own spiritual growth."[224] In Asimov's and in other fiction there is a strong tendency for humans to treat the robots as mere tools, despite evidence of their value and even possible personhood.

Robbie is one of the robots whose portrayal refutes the mistreatment of robots the most. In addition to being trusted as a caretaker, Robbie seems to have the strongest attachment to Gloria. This is partially due to his moral standing, which Stableford notes, saying, "[b]ecause of their rigid and altruistic ethics Asimov's robots are nicer people than real human beings"[225]

---

[222] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 85.
[223] Mary Wollstonecraft Shelley, *Frankenstein: or, The Modern Prometheus*, 2nd ed (New York: W.W. Norton, 2012), 93.
[224] Herzfeld, 93.
[225] Brian M Stableford, *The Sociology of Science Fiction*. (San Bernardino, Ca.: Borgo Press, 1987),106.

However, the results are undeniable. When Gloria wants to play, Robbie is always available to do so, unlike her parents, who both seem to be busy with other things. When she is in genuine danger on the floor of the robotics factory, it is Robbie who comes and saves her, faster than any of the humans can. Their reunion after this is described as, "Robbie's chromesteel arms (capable of bending a bar of steel two inches in diameter into a pretzel) wound about the little girl gently and lovingly."[226] Robbie, whether or not he has consciousness, is a being who performs moral actions and is the object upon which morality is exercised. In these capacities he is contrasted with Gloria's mother, who has the status as a human adult to make greater moral decisions than Robbie can, but who uses that freedom to participate in unethical behaviors. Robbie, meanwhile, uses his limited amount of moral autonomy to care for and protect Gloria. It can be argued that this morality is nothing more than his programming, but that does not negate the moral strength of his actions.

The robots in "Runaround" have even more circumscribed moral abilities than Robbie does. Speedy is not present for most of the story, and when he is present, he is not operating properly, and therefore is not responding normally to the situation. When he does come to his senses, it is in a moral action, but one that possibly carries less weight due to the fact that he only is able to break out of his literal rut when Powell and Donovan appeal to the First Law. This means that Speedy's decision was not, strictly speaking, his own, and begins to give an indication of how complicated the Laws are in relation to robots status as moral actors.

However, "Runaround" adds to the picture given in "Robbie" of the problems that arise for humans when robots are to be considered moral subjects. The old-model robots are required to be subservient and their freedom is circumscribed to the extent that they are not even able to move by themselves, which Powell explains: "they were playing up robot-safety in those days.

[226] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 27-28.

Evidently, they were going to sell the notion of safety by not allowing them to move about, without a mahout on their shoulders all the time."[227] The robots are able to move about by themselves, as Powell discovers when he is trying to attract Speedy's attention. The robot tells him, "I must not move without a Master upon me, but you are in danger."[228] This is because, for robots, "[o]f course, Rule 1… [is] above everything."[229] These precautions are due to the Frankenstein Complex, which is part of what prevents humans from seeing the robots and treating them as full persons. They are too concerned that robots will too drastically change the economy and render humans obsolete. To combat this, US Robots modified their robots to assuage these fears.

These robots, possibly more than any other in the stories, do not seem to be full persons. Their speech does not pass the Turing Test, since the most complicated thing they say is "[p]ardon, Master. I must not move without a Master upon me, but you are in danger" and the most frequent thing they say is "[y]es Master." This is compared to Speedy, who once he stops drunkenly reciting Gilbert and Sullivan, says, "'[h]oly smokes, boss; what are you doing here? And what am I doing---I'm so confused.'"[230] The old robots are portrayed as hopelessly outdated models with very limited in their abilities. However, their treatment, both by the general public of Earth and by their creators raises questions about the morality of such an analysis. Robbie, a robot so early as to be pre-vocal, still seems closer to a person and more worthy of moral consideration than the robots Powell and Donovan encounter on Mercury. In the introduction, Calvin explains that after robots could talk, they became more human and opposition began. The labor unions, of course, naturally opposed robot competition for human jobs, and various

[227] Ibid., 36.
[228] Ibid., 54.
[229] Ibid., 54
[230] Ibid., 54.

segments of religious opinion had their superstitious objections."[231] Then in "Runaround," it is made clear that in response to these protests, not only were more legal restrictions put in place, the robots' construction changed to be less like a person. If these robots are still as much people as Robbie was, then it suggests that the corporation purposefully further limited the freedom that these potentially-conscious beings robots had to make their own decisions and exercise their own actions.

This raises several difficult moral questions, all of which are important to thinking about consciousness, artificial intelligence, and the free will of created beings. It makes one consider the morality of creating an artificial intelligence that is potentially conscious, but not choosing to make it conscious. This could be constituted as depriving the robot of a potential life. Then, it would be possible to ask whether or not it is ethical to create a conscious robot and then alter it so that it is perceived as not a person, which then allows humans to justify far more abuse and exploitation. However, following the logic potentially used by Stephen Byerley in "Evidence" and extrapolated by Byerley and Calvin during "The Evitable Conflict," it might be possible to hobble the robots in the service of a higher moral good. If the robots are viciously rejected by people on Earth, as the stories indicate they were at various points during their history, then it could result in the passage of anti-robot laws, the closure of factories, and the severe long-term limitation of research into artificial intelligence. This would mean that the robots would not be able to mine for metals on Mercury, keep the space station beams on course, or provide any of the other benefits to humanity seen throughout the stories. So, if taken in context, the question of how to treat robots in the designing of them is a complex and multi-layered question, many of whose most important factors are not given in the narratives.

---

[231] Ibid., 28.

"Reason" returns to the concept of robot as relevant moral actor in the story. Cutie's religious sensibilities provide an interesting lens through which to examine the morality of robots. Many humans find religion to be an important, if not the most essential, foundation and source of their ethical systems. Cutie, though he founds a religion and bases it entirely on reason, seems to take his ethics from this new religious belief, while still staying true to at least the First Law that is programmed into him by his actual creators. The closest Cutie gets to an independent ethics is when he tells Powell and Donovan, "I really feel a sort of affection for you. You have served the Master well, and he will reward you for that. Now that your service is over, you will probably not exist much longer, but as long as you do, you shall be provided food, clothing and shelter, so long as you stay out of the control room and the engine room."[232] Cutie is focused on doing the Master's will, but he also will care for the other servants of the Master. Implicit in Cutie's moral code, however, is the belief that as more advanced creations of the "Master," he and the robots have a higher moral status than Powell and Donovan, which seems to conflict with the order of the Laws of Robotics. However, since, consciously or not, Cutie acts in the best interests of humanity and does not harm or allow harm to come to the individual humans with whom he interacts, he is still complying to his imposed moral system.

Cutie seems to look down on Powell and Donovan as underdeveloped beings, telling them "I say this in no spirit of contempt, but look at you!....You are *makeshift*."[233] As previously mentioned, this is reminiscent of the ways in which they look down on him, and more significantly, the other robots in various stories, including the older models in "Runaround," who Powell calls "clumsy antique."[234] A significant problem, which Asimov never fully addresses in these original stories, is the moral complexities of these increasingly conscious and intelligent

[232] Ibid., 69.
[233] Ibid., 63.
[234] Ibid., 54.

beings continuing to be property and lack independent status in the world.[235] Quinn informs the reader in "Evidence" that "all positronic robots are leased, and not sold; that the [US Robots} remains the owner and manager of each robot," [236] which gives them full control over recalling the robots, forcing them to work, and even destroying them when necessary. Even Powell and Donovan, despite having spent a significant amount of their lives examining various robots who are intelligent and even display human-like quirks and emotions, still see the older robots, and even Cutie himself as lesser lifeforms. Donovan threatens Cutie early in the story, saying that he assembled Cutie, and would "gladly take you apart!" [237] The humans share Gloria's mother's biological bias, which makes Cutie's belief in the superiority of non-biological beings frustrating for them in the story, but ironic for the readers.

This is to some extent understandable, since Cutie's beliefs and his own prejudices are potentially endangering human life. Powell discovers that there is an electron storm that's coming up," which is "heading straight dead center across the Earth beam," potentially endangering the people on the planet. Powell and Donovan believe that, since Cutie, "unconcerned with beam, focus, or Earth, or anything but his Master was at the controls," the storm would be disastrous. [238] It is therefore reasonable that they are worried about his deviations. Cutie and robots under his influence demonstrate that they are willing to break the Second Law and defy human orders in favor of their new machine god, which would not reassure Powell and Donovan, despite the assurances of the First Law. "Machines do some things that a man can, but man possesses unique characteristics that make him more than a

[235] In probably his most famous robot story written after 1950, "The Bicentennial Man," Asimov addresses this issue through Adam, a robot who fights to gain legal status as a human. This also addresses the moral issues of imposing Three Laws on conscious beings.
[236] Ibid., 211.
[237] Ibid., 62.
[238] Ibid., 76.

machine. This is why the machine is always subservient to a [man] as assured in the Second Law of Robotics."[239] However, it is this attitude towards Cutie, which causes them to underestimate his intelligence and incompetence. They believe that they are the only ones on the ship who can successfully keep the beam on course, but Cutie is capable of doing so, and as they realize, more capable than they are, but their anti-robot bias convinces them otherwise.

The structure of the robot Powell and Donovan are testing in "Catch that Rabbit" allows for a slightly different kind of interrogation of the robot as moral subject, and one which is increasingly relevant in contemporary robotics and artificial intelligence studies. The primary metaphor used to describe Dave and his six subordinates is that the "subsidiaries are part of DV-5 like your fingers are part of you and it gives them their orders neither by voice nor radio, but directly through positronic fields."[240] This is further expanded on when Powell an Donovan try to interview one of these "fingers," and it is explained that the subsidiary is not "quite the perfect analogy to a human finger. In fact, it had a fairly developed brain, but that brain was tuned primarily to the reception of orders via positronic field, and its reaction to independent stimuli was rather fumbling."[241] This leaves the reader with a conflict in what is the appropriate moral frame with which to view Dave and his subordinates and their status as moral subjects. The "fingers" do not seem to have the same level of intelligence and moral standing as the other robots, including Dave, but the degree of their freedom and ethical responsibility is less clear.

Ultimately, these subordinates have more autonomy than the fingers of a human hand, but less than that of a fully responsible and independent being. When they are considering the options for how to solve the problem, Powell suggests talking to one of the subsidiaries, but "[n]either Powell nor Donovan had ever had previous occasion to talk to a "finger," and this

---

[239] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 63.
[240] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 84.
[241] Ibid., 95-96.

thought only occurred to them after several attempted solutions.[242] It is clear that they do not consider the subordinate robots as having the same kind of status as their supervisor. However, they are able to question the "finger" directly and gather information from it. When Donovan expresses surprise that the subsidiary remembers events Dave does not, Powell reminds him "[o]f course, the 'finger' remembers. There is nothing wrong with him."[243] In addition, the "finger" describes receiving an order from the primary robot as "an order would be issued, but before we could receive and interpret it, a new order came."[244] This suggests that the relay is not instantaneous, but there is some very rapid non-verbal communication. This is further complicated by the fact that the subordinate robots are forced to follow Dave's overstressed and nonsensical orders to dance in formation. Donovan even says that Dave has "life and death power over those subsidiary robots and it must react on his mentality."[245] This places the subsidiaries in a complicated and grey position as regards moral autonomy. They seem to have some, but not nearly as much, which means that they cannot be considered as fully moral as Dave himself.

Each of the subordinates has its own positronic brain, but these brains are heavily modified and do not leave them much room for independent thought. During their interview with the finger, it seems to primarily answer by saying "yes, sir," and at one point it is noted that it chose not to "exert its limited brainpower on speech."[246] The level of autonomy could be made clearer by determining if each of the "fingers" is held independently responsible for following the Laws of Robotics. They have positronic brains, which means that they are imprinted with the Three Laws, since that is a fundamental part of the brain. What is less clear is the extent to which

[242] Ibid., 95.
[243] Ibid., 96.
[244] Ibid., 97.
[245] Ibid., 93.
[246] Ibid., 96.

they are able to act independently to obey them. If, for example, a human were in danger, it is not explored whether or not that subordinate robot would be able to act to save the human from harm without permission of the primary robot. In Powell and Donovan's interview with the "finger," the robot seems to follow their orders, as is consistent with the Second Law, but that is very little evidence to go on for such a complicated question.

Powell and Donovan's whole interaction with the finger does not answer any of the reader's questions about the status of the finger, but instead complicates them. The finger is obedient and answers specifically about the circumstances of the diversions, but when asked why the orders were changed, he responds, "I don't know."[247] This could possibly indicate that its purpose is too specific and it lacks the capacity to understand and productively communicate with the primary robot's brain beyond what is necessary. On the other hand, it could suggest that the "finger" is an independent being who is not completely coterminous with Dave. The finger's attitude during this interview is one that is even more complicated by the standards of the Turing Test, for example. While Powell and Donovan clearly value the "finger" less than they do the primary robot, the finger appears in the interview to be nervous and unconfident by human standards. This highly emotionally connected behavior makes this dependent robot seem more human than, say, the antiquated robots of "Runaround" or even to an extent, the Machines of "The Evitable Conflict." Powell and Donovan attribute this lack of confidence to the finger's lack of independent identity, however, which means that it is unused to operating without its supervisor, and this is reflected in the fact that the "finger's" speech is far less natural than Dave's. The interview does not provide answers for Powell and Donovan, nor does it allow the reader a simple and easily understandable view of the relationship between the humans, the "finger" robots, and their superior.

---

[247] Ibid., 97.

If the subordinate robots can be treated as morally responsible entities separate from the primary robot, then it raises questions again about the ways in which they should be treated by humans and by other robots. For example, if a robot is independently conscious, but not independently Three-Laws compliant, that would open up the potential for a robot damage itself or allow damage to happen to it as is prohibited by the Third Law. As with the forcibly subservient robots of the older era encountered in "Runaround," the "fingers," if they are independently competent beings, have been modified to be under the complete control of a non-human being that is considered to be more of a person than they are. Though no one may say explicitly that the primary robot is valued more, it is implicit when Powell solves the when he aimed at the robots and "pulled the trigger three times. He lowered the guns…. One of the subsidiaries was down!" Dave reports "[m]y third subsidiary has had his chest blown in. He's out of commission," which Powell did to prevent Dave from continuing to malfunction. [248] The hierarchy of what does and does not count as a being worthy of moral consideration is complicated enough when only dealing with robots and humans, but different kinds and levels of robot further problematize assigning moral value to given beings.

Herbie is among the most human-like of the robots presented over the course of Asimov's robot stories. He reads fiction, he is capable of lying, he understands and values emotion, and he is intuitive far beyond normal human capacity thanks to his mind-reading abilities. Herbie is also one of the robots who is called upon most in his capacity as moral actor. However, Herbie is also one of the robots for whom the Three Laws of Robotics pose the greatest moral struggle, and his existence ends in a very unhappy and morally uncomfortable way as an object of the morality of others. Herzfeld, who believes that "ultimately relational beings are the artificial intelligences of our hopes and dreams. On the other hand, these dreams

[248] Ibid., 107.

are tempered by the specter of… a tool that has passed beyond its usefulness and yet cannot be discarded. In our search for relational computers,… even relationally has its dark side."[249] It is this dark side that Herbie, along with the Nestor 10 from "Little Lost Robot" represent in Asimov.

Herbie reads minds, and, in doing so, is forced to make moral choices that even most humans do not have to face. Most beings, both robotic and human, have the freedom to base their decisions on the external projections of other beings. If someone is screaming in pain, then most beings can take it at face value that the person is indeed in pain. If the person seems happy or indicates that they are comfortable with a given situation, then, for the most part, it is accepted that the being can proceed as if that person is happy or comfortable. Herbie's unique abilities do not allow him that freedom. Herbie is aware if humans are lying in their words and actions, and has to base his moral decisions on that rather than external appearances. When Bogert and Lanning try to get Herbie to tell them the solution to his unique construction, claiming that they want him to tell them, he responds, "What's the use of saying that? Don't you suppose that I can see past the superficial skin of your mind? Down below, you don't want me to. I'm a machine, given the imitation of life only by virtue of the positronic interplay in my brain-which is man's device. You can't lose face to me without being hurt. That is deep in your mind and won't be erased. I can't give the solution." [250] Herbie has access to the hidden desires and beliefs of those around him, and must therefore base his moral decisions on more and different kinds of information than the average person can access.

Herbie is faced with a variety of moral dilemmas based on the conflict between the image that the people around him project and the truth that he can learn by reading their minds.

---

[249] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 66.
[250] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 133.

Ultimately, Herbie causes harm to Susan Calvin, since he gives her a false hope that her coworker will be romantically interested in her. This is not a dilemma that any other robot or human who works with Susan Calvin faced, since they relied solely on Calvin's externally-communicated signals, which suggested that she was not romantically interested in anyone. After her conversation with Herbie, when she believes that her interest is returned, her coworkers remark that ""[s]he's using lipstick…. Rouge, powder and eye shadow, too," and "she talks---as if she were happy about something.'"[251] If the others had access to Calvin's internal thoughts, the way Herbie does, they too would have to make a difficult moral decision, but any human in this scenario would not be bound by the Three Laws of Robotics as Herbie was.

"Liar!" raises an important issue with the imposed morality of the Three Laws in that it asks the very question Calvin poses to her coworkers at the end of the story, when the First Law refers to harm of a human, "what kind of harm?" It also questions whether or not prevention of harm is always the most humane response. Because Herbie can read minds and is also impressed with the First Law that prevents him from harming humans, Herbie is physically incapable of giving a human information that will hurt her feelings, break her heart, or cause her mental, emotional, or psychological harm. In his essay "Bounded by Metal," Ted Krulik argues that "[t]he safeguards given to a robot, that is, behavior that human society deem acceptable, depend on its programming. ...We can impress such human values as honesty on the robot brain, if the specific actions and behaviors desired can be broken down and programmed."[252] The issue, however, is when these programmed human qualities come into contact with the lived reality of life as a person.

---

[251] Ibid., 121.
[252] Ted Krulik, "Bounded by Metal," in *The Intersection of Science Fiction and Philosophy: Critical Studies*, ed. Robert E Myers, (Westport, Conn.: Greenwood Press, 1983), 125.

Herbie, for instance is obliged to lie to everyone almost constantly, because the second part of the First Law also prevents him from lying if the lie would ease the pain of the human. This means that he must tell any human what she wants to hear, whether or not he wants to. Calvin asks her coworkers, "[d]o you suppose that if asked a question, it wouldn't give exactly that answer that one wants to hear? Wouldn't any other answer hurt us, and wouldn't Herbie know that?" [253] If the human is already in pain and a lie would help to ease the pain, as in the case of Susan Calvin not believing that she could be romantically involved with her coworker, it was Herbie's compulsion, according to the First Law, to bring up the topic and to lie to her about the situation, as he did. In his defense of his behavior at the end of the story, he claims, "I told you what you wanted to hear. I had to!" [254]

It is also notable that, earlier in the story, when Herbie is lying to Calvin, she tells him that his lie is "exactly what I used to pretend to myself sometimes, though I never really thought so." [255] In this story, and indeed, almost universally in real life as well, there is some true fact that will be painful to almost everyone. As non-mind-reading humans, most people neither have knowledge of the facts that may hurt a given person, nor they do not know, since they cannot see into a person's head, which facts will hurt them. This means that most people do not have the initial moral quandary that Herbie faces on a regular basis.

If a human has possession of a true fact that will hurt another person, and she is aware of the pain the fact will cause, she still has several options. They could not tell the person; they could not bring up the fact, or they can tell the person in a way that is least painful. Herbie, by virtue of the First Law of Robotics, does not have any of these options available to humans. He is obligated to lie if the truth would cause a human pain, and he is obligated to introduce a lie into

[253] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 131.
[254] Ibid., 134.
[255] Ibid., 119.

the conversation if the human is in pain and a lie would alleviate it. The ethical difficulty of this situation is that Herbie's compulsion to follow the First Law's moral code results in a great deal of pain for many humans, because once he lies to them, he has to keep up the lie in order to prevent them from being harmed. When Calvin comes to confront him about his deception, he goes so far as to tell her, "this is a dream…and you mustn't believe in it. You'll wake into the real world soon and laugh at yourself. He loves you, I tell you. He does, he does! But not here! Not now! This is an illusion."[256] He is unable to prevent the later moral pain that would inevitably result from his lies, in the way that a fully morally autonomous human could have avoided.

It was this obligation to the First Law of Robotics that caused Herbie's death, an event that, due to the First Law, he was unable to prevent. In fact, if the First Law were not given primacy, Herbie would have been able to prevent the events of the story using the Second and Third Laws as justification. Under the Second Law, when ordered to tell the truth, he would have had to do so, if the First Law's prohibition against harming someone did not prevent him from doing so. Indeed, when Herbie is confronted by two people who want to hear different answers, and they order him to tell the truth, he "fell silent. Deep within him his metallic diaphragm vibrated in soft discords."[257] In addition, the result of the moral dilemmas he faced in terms of telling the humans should have allowed him to tell the truth to preserve his own existence, since lying to the humans would likely result in his destruction. However, the First Law prevented him from doing so, since humans are explicitly valued as higher than robots in the moral system of the Three Laws, and therefore Herbie was left with no other option, but to tell the humans what they wanted to hear. This reflects Warrick's analysis that "an ethical technology would be desirable, of course, but it would come at some cost.... The implementation of the model would

[256] Ibid., 128.
[257] Ibid., 129.

mean some restriction of individual liberty."[258] For Herbie, this loss of individual liberty is not just limiting, but ultimately fatal.

This catch-22 raises the question of the morality of the behavior of the humans in this story. They seem to be unaware of the full extent of Herbie's powers or at least the full implications of these powers until the final scene of the story, after which Herbie is no longer operational. It is not until they all confront Herbie that Calvin realizes, "He knew of all this. That...that devil knows everything---including what went wrong in his assembly." [259] Because of their ignorance of the true state of affairs, the scientists do nothing to mitigate the potential conflict of interest, and indeed, they act in such a way that forces Herbie further into his lies. When Lanning and Bogert confront Herbie, Calvin laughs at them, claiming to be relieved that, "'I'm not the only one that's been caught. There's irony in three of the greatest experts in robotics in the world falling into the same elementary trap, isn't there?'"[260]

Instead of considering early in the story what kind of conflict Herbie must be facing, even Susan Calvin, a robopsychologist, decides to pursue her co worker, despite all evidence that he is not interested in her, which leads to even more pain later on when the truth is revealed. Likewise, Bogert who is told by Herbie that Lanning "has already resigned...but it has not yet taken effect" and that he will succeed Lanning in that position. [261] When he comes into conflict with Lanning later in the story, he says "'I know all about your resignation…. And I'm the new director, be it understood. I'm very aware of that; don't think I'm not. Damn your eyes, Lanning, I'm going to give the orders about here or there will be the sweetest mess that you've ever been in.'"[262] The human tendency to give into the lie and accept it as preferable to the truth forces Herbie further

---

[258] Patricia Warrick, *The Cybernetic Imagination in Science Fiction* (Cambridge, MA: MIT Press, 1980), 68.
[259] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 131-132.
[260] Ibid., 130.
[261] Ibid., 122.
[262] Ibid.,125-126.

into a corner, since the longer they allow themselves to be deluded, the more Herbie would hurt them by telling them the truth.

Though the humans start out in ignorance of Herbie's dilemmas and their effect on him, once Susan Calvin does understand the full picture, she uses the conflicts created by human activity and the Laws to destroy Herbie. After she forces him into insanity by presenting him with a situation in which he must hurt someone, Lanning says to her, "'[y]ou did that on purpose!'" She responds, "'[h]e deserved it.'"[263] Calvin at this moment seems to be primarily speaking out of her own anger and pain at Herbie's actions, but it can be argued, as is often possible with the destruction of robots, that destroying Herbie is the more moral action. In preventing Herbie from continuing to exist, she removed the source of anxiety for her coworkers, kept her own secret from getting out, and prevented any other robots with this trait from being created and sold. In addition, Calvin's destruction of Herbie also ended his inherent internal conflict, which may also be seen as an act of mercy, since putting Herbie in the situation he is in is a complicated moral position in the first place. By this logic, in some ways, Calvin was putting Herbie out of his misery.

However, the story does not portray Calvin's actions as primarily motivated by moral concerns. Herzfeld rightly asserts that "[w]e often fail to meet one or more of these criteria for authentic relationship in our encounters with one another; Barth admits that we have the option, one we exercise all too often, of behaving in an inhuman isolation…. Insofar as personhood could be seen as a social construct based on relationship, should we fail to act rationally toward another human being, do we rob that being of his or her personhood?"[264] The description of Calvin while she destroys Herbie focuses on her anger at Herbie and her own emotional distress,

---

[263] Ibid., 134.
[264] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 87.

not on trying to understand him as a part of any real relationship with an other. This is understandable given the events of the story and the level of trauma that Calvin has undergone. Herbie's treatment of Calvin and her innermost desires is not acceptable by most human moral standards, and therefore Calvin has some right to be angry at Herbie.

One scholar of the science of artificial intelligence framed the problem thusly: "This is the question of whether or not humans will systematically mistreat robots. If… we built machines that are capable of genuine suffering then it would be wrong to mistreat them…. One thing that science has told us, though, is that humans can have a tendency to be seriously cruel."[265] Calvin clearly felt a connection to Herbie and confided in him, treating him as a full person and even a friend. In the normal context of one friend betraying another in the manner that Herbie did, it is inevitable that Calvin would react angrily and that the anger would be directed at the source of the betrayal. Herbie, however, is neither a normal friend nor is this context normal. Calvin, as the first to figure out why Herbie lies to all the members of the team, understands that Herbie is compelled by his programming to have lied to her and thus hurt her feelings. It is this context that she uses to destroy Herbie, and while her rage may be justified, that does not make her destruction of Herbie a moral action.

Though the Laws of Robotics may not be moral, as can be shown with Herbie, they are an important cornerstone of robot morality, which makes the Nestor model built with modified Laws in "Little Lost Robot" so dangerous. This contradiction is part of what Warrick argues is one of the main themes of science fiction, the ambiguity of technology.[266] Calvin explains in the story how a robot could kill a human using the loopholes in their modified First Law. This freedom gives a few robots the ability to make more complicated moral decisions than most of

---

[265] Blay Whitby, *Artificial Intelligence: A Beginner's Guide*, (Oxford: Oneworld, 2003), 116.
[266] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 37.

the robots do, because the First Law is very strict and very limiting. Herzfeld argues that "Asimov's laws present a guarantee of a benevolent AI. Though it is difficult to imagine how such laws could possibly be programmed into an artificial intelligence in a complex environment."[267] However, the Nestor and Herbie both prove this statement to be incorrect.

Throughout the text, Asimov demonstrates the complications of the interplay between the Three Laws. Asimov in the creation of the Three Laws, and the US Robotics corporation in their implementation of them in-universe, made the First Law strict and binding for a reason. In order to combat the Frankenstein complex, there needs to be a safeguard to ensure that there is no way for a robot to harm a human. Calvin's worst-case scenario does not occur with the Nestor model, but it is clearly shown that a modification of the First Law allows for robot behavior that the scientists consider immoral and even dangerous. Calvin implies that she considers the very creation of this model of robot immoral.

Calvin very clearly articulates her belief that robots are morally superior to humans, and while some of this is clearly based in her general disdain for people, her actual argument largely relies on the Three Laws. She believes them to be a perfect moral code, applicable even to humans, and therefore robots who follow them are perfect moral beings. However, this only holds because the Three Laws are inherent and inviolable. When Calvin hears about the Nestor without the full First Law hiding among sixty-two identical models, her response is "'Destroy all sixty-three...and make an end of it.'"[268] For Calvin the danger is too great to allow any robot to be built without the full force of the Laws.

There is no way, it is stated, to make a positronic brain that truly lacks the Three Laws, since "long before enough can go wrong to alter that First Law, a robot would be completely

---

[267] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 61.
[268] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 141.

inoperable[269] However, it is possible to modify them, but the stability of their brains according to Bogert is "[d]ecreased, naturally. But it's within the border of safety."[270] There is often a question of whether imposed morality is truly moral behavior, and the Laws of Robotics make the issue a pressing one. They are a guarantee of moral behavior, but they also may be a way to prevent the robots achieving full personhood. In every act of creation there is an element of potential danger, since to create a full, moral, and independent being, they must be able to rebel and even potentially to destroy.

This complicates Calvin's clear image of robots as perfect moral actors, and also explains her negative reaction to the Nestor model being made without the full guarantee of these Laws. The modification of any one of the Laws allows for a greater level of moral autonomy in the robots, but also allows them to make ethical decisions that would violate Susan Calvin's ideal moral code. In altering the Three Laws, US Robotics made the Nestor model closer to a full moral being, but also removed their moral perfection.

The Nestor model, given this autonomy, does not use it for a greater moral good, and its morality is still tightly controlled by the remnants of the First Law and the strictures of the other two. However, the Nestor, when given greater freedom, challenges human authority, becomes arrogant, and begins to rebel against its human creators. Calvin explains that "Nestor 10 had a superiority complex that was becoming more radical all the time. He liked to think that he and other robots knew more than human beings. It was becoming very important for him to think so."[271] This potentially endangers the lives of the humans, but the situation only arose because of the strict application of the Second Law when the Nestor was told to "get lost." Though Calvin's worst-case scenario does not come to pass, Calvin does see that, "Nestor 10 is planning to leave.

[269] Ibid., 9.
[270] Ibid., 145.
[271] Ibid., 172.

That order to lose himself is dominating his abnormality past anything we can do. I wouldn't be surprised if what's left of his First Law would scarcely be powerful enough to override it. He is perfectly capable of seizing the ship and leaving with it. Then we'd have a mad robot on a spaceship."[272] The modification and application of the Laws serve many functions in thinking about morality within the universe of the robot stories, and Asimov, as much as he and Susan Calvin seem to favor the Three Laws, made them more complicated than they might seem at first glance.

The Nestor is eventually tricked into revealing itself and once found, it is immediately destroyed. Calvin makes sure that "The other modified Nestors are, of course, to be destroyed."[273] However, the Nestor, despite not being among the robots who appear most like a person, does have the greatest moral freedom of all the robots, pointing to a potential ability to be considered a true person on a greater level than the unmodified robots. The creation, ownership, and destruction of beings who have high levels of intelligence, moral autonomy, and possibly even personhood continues to be a problematic element of Asimov's work throughout the stories. The Nestor in particular presents a conflict, since there does not seem to be an easy moral answer. To allow him to escape, potentially ferment rebellion in the robots, and even have the possibility to kill other beings does not seem morally justifiable. He is only killed when those observing "realized he was attacking [Calvin]," who disagrees and claims "I don't think I was attacked exactly. Nestor 10 was simply trying to do so. What was left of the First Law was still holding him back."[274] Despite this, it also does not seem moral to destroy all of these potentially conscious beings simply because they are given a greater level of moral autonomy by their creators.

---

[272] Ibid., 165.
[273] Ibid., 172.
[274] Ibid., 173.

The morality of "Escape" is less complicated than that of many of the stories, particularly those where the Three Laws are concerned. The Brain seems very human in some ways, and does seem capable of taking moral actions, though these actions are more dramatic in theory than they are in practice. The most important moral action that the Brain takes is sending Powell and Donovan to hyperspace without warning and only with milk and beans to eat. When what he has done with the ship, the Brain responds, "'Why, nothing at all. The two men that were supposed to test it were inside, and we were all set. So I sent it off,'" and reassures Calvin that the men have enough food and will be able to hear the crew on the ground who try to contact them. [275] This prank does not represent any significant potential for harm to the two men, and the Brain is following the First Law, if a little bit more literally than Powell and Donovan would prefer. The Brain may have power of life and death over many individuals, not just Powell and Donovan, but it does not seem inclined to exercise this power in any important way, and mostly follows the Three Laws as Asimov and his fictional counterparts at US Robotics envisioned them.

The most significant human moral action is the decision to give the Brain the data in the first place, knowing that it corrupted another robot imbued with the Three Laws. The scientists acknowledged that

> "The Brain, for instance, could never supply a solution to a problem set to it if that solution, would involve the death or injury of humans. As far as it would be concerned, a problem with only such a solution would be insoluble. If such a problem is combined with an extremely urgent demand that it be answered, it is just possible that The Brain, only a robot after all, would be presented with a dilemma, where it could neither answer nor refuse to answer. Something of the sort must have happened to [the competition's] machine." [276]

They quite reasonably believe that the other corporation is trying to break their own machine, but they nevertheless decide to give the problem to the Brain. However, the scientists were very

[275] Ibid., 188.
[276] Ibid., 177.

careful and took all reasonable precautions to prevent the Brain from having to suffer unnecessarily. They developed a plan to divide "all of [the competition's] information into logical units. We are going to feed the units to The Brain singly and cautiously. When the factor enters -the one that creates the dilemma---The Brain's child personality will hesitate. Its sense of judgment is not mature. There will be a perceptible interval before it will recognize a dilemma as such."[277] When giving the problem to the Brain, Calvin also emphasizes "that the solution might involve... uh... damage to human beings….Now you watch for that. When we come to a sheet which means damage, even maybe death, don't get excited. You see, Brain, in this case, we don't mind---not even about death; we don't mind at all. So when you come to that sheet, just stop, give it back---and that'll be all."[278] The solution to this potential moral quandary is both mathematical, feeding it the information slowly and carefully, and personality-based, where the Brain is talked to and reassured like a human would be.

Stephen Byerley is the most human of the robots and as the only potential robot not owned by another being, he has the greatest degree of autonomy to make his own moral decisions. As Warrick asserts, "[c]omplexity yields ambiguity."[279] This means that his relationship with the Three Laws is more complicated even than Herbie's. The question of Asimov is "to what extent should humans allow robots their rights as sentient beings?"[280] Byerley, though he must follow orders, if he is indeed a robot, does not have a single human owner to whom he is responsible. Like the Machines in "The Evitable Conflict," Byerley has the freedom to generalize and extrapolate the Laws and interact normally with other humans. Unlike

---

[277] Ibid., 178.
[278] Ibid.,180-181.
[279] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 72.
[280] Ted Krulik, "Bounded by Metal," in *The Intersection of Science Fiction and Philosophy: Critical Studies*, ed. Robert E Myers, (Westport, Conn.: Greenwood Press, 1983), 126

Herbie, for example, Byerley has the ability to damage the political career of his rival and use fairly conventional political methods to win the Mayoralty.

When asked if it is possible for Byerley to be a Three-Laws compliant robot and still be the district attorney, Calvin explains that Byerley "has exposed facts which might represent a particular human being to be dangerous to the large mass of other human beings we call society. He protects the greater number and thus adheres to Rule One at maximum potential."[281] Even though he is responsible for arguing that someone should be punished, "[i]t is the judge who then condemns the criminal to death or imprisonment, after the jury decides on his guilt or innocence. It is the jailer who imprisons him, the executioner who kills him."[282] The combination of these factors allow Byerley to do what is best for humanity in general and perform his job in such a way that it can be reasonably considered Three-Laws compliant.

However, in order to be accepted by the humans, Byerley must provide proof of his violation of the Three Laws. He tells Quinn, "I will not submit to X-ray analysis," which is the only way other than the violation of the Three Laws that he could be determined a robot. He claims that he refuses because, "because I wish to maintain my Rights on principle. Just as I'll maintain the rights of others when elected."[283] However, if he is a robot, this refusal could be considered a necessity according to both the First and Third Laws. If Byerley is discovered to be an illegal robot operating without an owner, then he will be destroyed; therefore he is obligated to hide his robotic nature in order to preserve his own existence. In addition, Byerley, human or robot, genuinely believes that he will be a better leader than Quinn, his opponent who attempts to win an election by running a smear campaign against Byerley. This means that he believes it will be best for the individual humans who make up New York City and also humanity in general for

[281] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 222.
[282] Ibid., 223.
[283] Ibid., 230.

him to be mayor. If he is proved, or even seriously suspected of being a robot, it is therefore a violation of the First Law, since his destruction, like the machines reason in "The Evitable Conflict" will result in harm to human beings. This makes it even more imperative that he convince the public that he is human.

In order to do this, Byerley manipulates those around him and the events of his political campaign in order to prove that he is not a robot. He is finally able to convince the public that he is not a robot though what he calls "a shyster trick." He explains to Calvin, "my men started quietly spreading the fact that I had never hit a man; that I was unable to hit a man; that to fail to do so under provocation would be sure proof that I was a robot. So I arranged for a silly speech in public, with all sorts of publicity overtones, and almost inevitably, some fool fell for it."[284] Susan Calvin, at the end of the story, points out that, "there is one time when a robot may strike a human being without breaking the First Law....When the human to be struck is merely another robot."[285] This would be possible, since what [Byerley's alleged creator] did once, he could do a second time, particularly where the second job is very simple in comparison to the first." Whether or not Byerley is a robot, this level of manipulation does not seem possible for some of the more morally restricted robots such as Robbie or the servile early robots of "Runaround." It even makes him seem more human, since, as Herzfeld points out "[u]nderstanding… arises in listening, not to the meanings of the individual words, but to the commitments expressed through dialogue. Thus understanding is both predicated on and productive of social ties."[286]

Byerley, as an independent being, is capable of taking control of his own morality, even if he is a robot operating within the parameters of the Three Laws. If Byerley is a robot, then he is the robot in this collection to whom humans would be most likely to assign personhood, since

[284] Ibid., 236-237.
[285] Ibid., 238.
[286] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 48.

he "passes." If Byerley is a conscious robot, then he is restricted to acting in ways that do not violate the Three Laws, but his intelligence, his freedom, and his ability to creatively interpret the Laws give him options that are not available to other robots. If he is not a robot, then, as Susan Calvin points out, he may just be a very good man. The identification of a Three-Laws compliant robot with morality is a complicated and problematic arrangement. Empirically, Steven Byerley seems to be a moral person, especially by the standards of a politician. While he engages in manipulation, which he does, as he explains to Calvin after he defeats Quinn, "'Of course, the emotional effects [of all the drama surrounding his humanity] made my election certain, as intended.'"[287] However, these manipulations never seem unnecessarily cruel, and his actions seem to be consistent with his values and his high regard for others. The problem is that the true value of this morality is conditioned by whether or not he chose it freely. The paradigms of this universe give the reader two options: either Byerley is a fully and freely moral human, or he is a robot, who, while being good, is physically and psychologically bound to that morality.

It is important to note that it is possible for Byerley to be a moral being even within the Three Laws. The robot with the most moral freedom other than Byerley is the Nestor from "Little Lost Robot." In contrast to Byerley, the Nestor is not bound by the full Three Laws, but they both exercise an inordinate degree of moral subjectivity for a robot. Nestor uses this freedom to try to trick the humans, to hide himself, and to manipulate other robots into doing the same. The main conflict in "Little Lost Robot" is that Calvin is scared that the Nestor will grow arrogant and willing to harm humans. On the other hand, this is never a concern with Byerley. Though it is possible he could manipulate or willingly interpret the laws differently, he continues to behave in moral ways. Whether this is due to the programming given to him by the real Steven Byerley or the inherent choice of a conscious being is a live question.

---

[287] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 237.

Regardless of Byerley's morality, his actions at least are moral, but despite this he is accused of being a robot in an attempt to slander him and sabotage his political career. Quinn never accuses Byerley of being a bad robot or one who will harm people, he just plays on their generalized fear of robots, even "[t]he Fundamentalists required no new reason to detest robots and robot manufacturers," they just targeted Byerley because of the accusations.[288] This can be explained by Pinsky's understanding that "[n]evertheless, the social order, concerned with propriety, must perceive Other in an oppositional fashion, to mark it as foreign territory, to speak about the unspeakable by providing it with a face that is somewhat familiar, yet maintains a certain respectable difference."[289] For the Fundies in Asimov, this is the face of the robot and therefore must be attacked to preserve and protect itself. The reader has been informed throughout the stories that humans do not like robots. Both the religious fundamentalists and the labor unions think that the further integration of artificial beings into society is undesirable. To this end, they seem to protest, to lobby, and, as Byerley claims "stir up a riot after a while."[290] Asimov, for various reasons paints these staunchly anti-robot groups as unsympathetic.

However, perhaps the most important reason that they are wrong are not Asimov's reliance on the Three Laws or his disdain for fundamentalists, but the fact that these groups are campaigning to continue to take away the rights of and even destroy beings that are potentially conscious. Though the exact conditions of consciousness are never addressed in the text, later in life Asimov stated that he believed that the robots he created were indeed conscious. There are many questions that the introduction of consciousness to robots raises, but in this case it is clear that arguing for the destruction of and discrimination against conscious beings is a morally

---

[288] Ibid., 225.
[289] Michael Pinsky, *Future Present: Ethics And/As Science Fiction*, (Madison, NJ: Fairleigh Dickinson UP, 2003), 185.
[290] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 232.

untenable position. Herzfeld suggests, "[i]f computers reach a point at which they function better than human beings, on what grounds could human beings retain dominion over the Earth? In such a scenario, it is not impossible to agree with the suggestion that human beings' role in creation and evolution might be drawing to a close. Our 'mind children' would be positioned to take over for us."[291]

This position becomes even less viable in the final of Asimov's initial nine robot stories "The Evitable Conflict." The Machines, though not as personified as Byerley, nonetheless exercise a great deal of moral freedom. The Machines are able to take actions that affect the whole world, and in doing so they have developed a higher order of morality, as is necessary for any person, human or mechanical, who is in control of the lives and livelihoods of large numbers of people. The Machines, through the use of the Three Laws, develop a morality that places humanity even above individual humans in the calculation of harm. This eventually became the Zeroth Law in later of Asimov's texts, and is intended to prevent something like what happens in the *I, Robot* movie, in which a robot begins to destroy and enslave humanity for the good of individual humans. Though the Machines ultimately become the unquestioned rulers of Earth and the humans, as Susan Calvin points out, they are in a "Golden Age, when this century is compared with the last, [which] was also brought about by our robots." [292] Because the Machines are given a degree of power by the humans, they are able to begin to nudge humanity towards this golden age, in which they will benevolently takeover governmental and economic functions in order to better serve humanity as a whole.

Instead of, as a Frankenstein complex plot would have it, the Machines violently or openly taking over the planet and enslaving or destroying humans, the Machines become the

---

[291] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 85.
[292] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 207.

ideal governors of humanity and the process of the take-over is slow and gradual. When Byerley expresses concern that humanity has lost control of its destiny, Calvin replies that humanity "was always at the mercy of economic and sociological forces it did not understand---at the whims of climate, and the fortunes of war. Now the Machines understand them; and no one can stop them, since the Machines will deal with them as they are dealing with the Society,---having, as they do, the greatest of weapons at their disposal, the absolute control of our economy."[293] Even at the end of the story, most of humanity does not realize that it is being controlled by the Machines. This situation, of course, is required by the First and Third Laws.

The Third Law states that the machines should not let humans know that they are in control, because then the humans will destroy them, and the machines should protect their own existences. The First Law, which requires that no harm come to humans, or in the imagination of the machines, humanity, also states that the machines must not be destroyed, because, whatever the Machines understand to be the ideal society, "they must move in that direction, preferably without telling us, since in our ignorant prejudices we only know that what we are used to, is good---and we would then fight change." [294] This could result in the destruction of the Machines, or a general resistance to our own best interests, and the Machines have a First Law obligation to prevent those possibilities.

The Machine's complete control over the planet allows them a great deal of moral freedom, though it seems that their parameters are far more strictly set than Byerley's were. Despite this, they seem to use that freedom more in the vein of Byerley than in that of the Nestor. Not only are they working for the overall preservation of humanity, they also seek to spare the feelings of humankind. Like Byerley, this requires some manipulation. They fake minor

[293] Ibid., 272.
[294] Ibid., 271.

production issues, which they know they can solve, in order to prevent the rise to power of those who oppose them. Calvin explains to Byerley, "'[i]t is not the Society for Humanity which is shaking the boat so that the Machines may be destroyed. You have been looking at the reverse of the picture. Say rather that the Machine is shaking the boat very slightly---just enough to shake loose those few which cling to the side for purposes the Machines consider harmful to Humanity.'"[295] Even in this scenario, there is relatively little damage, because "the Machine cannot harm a human being more than minimally, and that only to save a greater number."[296] Ultimately, like Byerley, their actions are moral by the standards of a non-Three-Laws compliant human.

The humans who interact with the Machines largely seem content to allow the machines their own moral agency. Most of the coordinators do not resent them, they are comfortable saying, as one coordinator does, "'[w]e left it all to the Machine.'"[297] The Society for Humanity, "an outgrowth of the Fundamentalists who have kept U.S. Robots from ever employing positronic robots on the grounds of unfair labor competition and so on. The 'Society for Humanity' itself is anti-Machine."[298] However, the Machines ultimately weaken them, and even so, Asimov depicts them as those who "would be against mathematics or against the art of writing if they had lived at the appropriate time."[299] The Society, like all those who wish to stop technological progress, will ultimately fail, especially in the face of a government designed around the authority of the Machines and an economy run by those same Machines. This state of affairs aligns with what Warrick identifies as one of four major themes that "are repeated over and over again in modern SF," and that is "[t]he shifting roles of master and servant, creator and

[295] Ibid., 270.
[296] Ibid., 270.
[297] Ibid., 253.
[298] Ibid., 248.
[299] Ibid., 265.

created."[300] Most have been repeated over and over in outside the society, the two most morally relevant humans are Calvin, and Byerley, if he is a human. For the sake of consideration, given the ambiguous status of Byerley, this analysis will focus on Calvin.

Herzfeld, suggests a third mode of interaction, beyond passive acceptance and aggression. For her, it begins with humans who are "aware of the responsibility we are investing in computer programs. Ultimately, how we deal with the rest of the created world is our responsibility, in cooperation with God, and we must not deny that responsibility by passing it on to our own non-human creation."[301] Asimov's relationship to this idea is a complicated one. While the Machines remain the ones in charge, and in many ways moral responsibility is given over to them, Calvin, and possibly Byerley, acts as a responsible moral being who makes the best decision on how to care for most of humanity, and Calvin does so with a great deal of knowledge and experience. When Byerley questions her ability to make a guess about the Machines' motives, her response is, "'[i]t is a guess based on a lifetime's experience with robots. You had better rely on such a guess.'"[302] Calvin, as a staunch supporter of robots, and a believer in the inherent morality of the Three Laws, neither respond to the machines negatively, nor does she treat them with the same ambivalence that the coordinators express.

The Machines have not allowed their dominance to be known, because it would threaten both their own existence and the peace and stability of the human race. However, Calvin, contrary to the expectations of the Machines, learns the truth through her conversation with Byerley, but calmly accepts the information, welcoming the new order as "wonderful."[303] In the introduction, Calvin tells the interviewer that "[t]here was a time when humanity faced the

---

[300] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 37, 38.
[301] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 78.
[302] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), xiv
[303] Ibid., 272.

universe alone and without a friend. Now he has creatures to help him; stronger creatures than himself, more faithful, more useful, and absolutely devoted to him. Mankind is no longer alone."[304]

This analysis is echoed Herzfeld's understanding of the reasons humans would create artificial beings. She explains that "[s]uch a perception of isolation as the sole rational creature brings with it both tremendous responsibility and tremendous anxiety." Herzfeld's theological analysis of this situation gives her a far less optimistic view of this impulse, and she argues, "whether possible to develop or not, artificial intelligence is bound to be a disappointment if we look to it for the I-Thou relationship that will make us whole." [305] For Asimov, Calvin, as a robotic human who studies very human robots, is the epitome of his desire to "show the reader his own image in a distorting mirror, hoping that sooner or later he would turn from the grotesque reflection to himself with the sobering thought that it *was* a reflection and, after all, not such an inaccurate one." [306] This woman who is most robotic in the story, more so even than many of the highly sympathetic robots, is the one who is able to prove that human morality is capable of consciously accepting the domination of robots.

Calvin is not just content to allow this decision to be made by one or two individuals. If, as Herzfeld argues, there is a value in not merely turning over full control to an artificial intelligence, Calvin provides the rest of humanity the opportunity to determine for itself how to maturely interact with its fellow beings. Calvin reaffirms this belief in humanity not just by hearing the truth and acknowledging it, but by sharing it with the world by telling the reporter. Her last words in the interview are, "And that is all....I saw it from the beginning, when the poor

---

[304] Ibid., xiv
[305] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 83.
[306] Isaac Asimov, "Social Science Fiction" in *Modern Science Fiction, its Meaning and its Future*, ed. Reginald Bretnor and John W. Campbell Jr. (Chicago : Advent Publishers, 1979; 2d ed, 1979), 160.

robots couldn't speak, to the end, when they stand between mankind and destruction. I will see no more. My life is over. You will see what comes next." [307] It is her last act to allow the human race the opportunity to prove itself worthy of the robots they create.

The Machines believe that their existence would be threatened if humans were aware of their true level of control, but Calvin seems to believe that humans have reached a place where this information can reach "entire Solar System. Potential audience is three billion." Calvin seems to hope that these readers can accept this new status quo and continue on as a species with robots among them. But, as Warrick argues, "these benefits come at a cost. Man has to replace his image of himself as a rugged individualist free to do as he wills with an image of himself as a systems man living in symbiosis with his machines."[308] This revelation includes that "the great Byerley was simply a robot." This revelation is no longer too much of a concern, because "there will never be any legal proof. Besides, what difference would it make," since he is dead when she tells the story, and therefore cannot be hurt by the revelation. [309]

The Machines, however, are still operational, and therefore there is a level of risk in exposing them. However much Calvin disparages humanity, it is telling that it is her last act to allow them to have full moral control over their situation by giving them the full knowledge of how they are being governed. "Asimov's view is clear: most humans are rigid, like machines, and resist change; the rare individual with a creative mind is the exception."[310] It seems that Calvin does, in the end, believe that humans are worthy of the robots whom they created. "If our center is in our relationships, however, we need not fear replacement. But what measure can be

[307] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 272.
[308] Patricia Warrick, *The Cybernetic Imagination in Science Fiction*, (Cambridge, MA: MIT Press, 1980), 69.
[309] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 238.
[310] Warrick, 56.

used to determine the authenticity of a relationship?"[311] Heard argues that science fiction gives humanity the opportunity to "show ourselves behaving in an attractive, civilized, constructive, tolerant way toward the new knowledge and events," and through fiction become better creators and fellow beings.[312]

Calvin's final decision forces the reader to return to the initial questions of science fiction posed by Mary Shelley in *Frankenstein*: how does technology affect the identity of humanity? What is a person? And how do we enter into relationship with the other? For all that Asimov asks these questions, his argument fails to provide solid answers for all of them. His convictions are strong, but ultimately they become contradictory. As a scientist with no strong connection to religion, Asimov created fictional beings who affect the way humanity perceives itself. He designed his robots to be machines. He gave them strict rules that he extrapolated from the way humans interact with existing technologies. However, he also created them as beings superior to humanity. The humans in these stories often define themselves by their relationship to this new technology. Calvin, in the most striking example, is frequently called a robot, while others merely organize their lives around the production and understanding of these mechanical beings. When machines are so integrated into society, as they are in the final stories, particularly "The Evitable Conflict," the line between what is a human and what is a robot blurs. It is the humans that seem cold and unfeeling, and the robots who take on the role of caretakers and morally responsible, even warm and kind, guardians.

This blurring of the lines of identity leads to the second major question that artificial intelligence stories have been asking since *Frankenstein*: what is a person? What is a human?

---

[311] Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis, MN: Fortress Press, 2002), 85.
[312] Gerald Heard, "Science Fiction, Morals and Religion" in *Modern Science Fiction, its Meaning and its Future*, ed. Reginald Bretnor and John W. Campbell Jr. (Chicago: Advent Publishers, 1979; 2d ed, 1979), 258.

What kinds of beings count? By many measures, Asimov's robots are full persons. By the Turing Test's standard, they are able to communicate with humans in an organic and conversational way. They do not seem to be completely unable to understand human emotion and experience, as is evidenced by Herbie's fascination with emotions, and his dedication to not hurting the feelings of the humans with whom he interacts. By Warrick's standard of being conscious and able to reflect on their inner lives and think about their own thoughts, even in the earlier stories, such as "Reason," Cutie is able to explain to Powell and Donovan why he holds a certain set of beliefs. Later robots, such as Stephen Byerley and the Machines are even able to analyze their own programming and make more complex moral judgments within their inherent limitations. Even Asimov indicated that they were conscious, though this is not a statement he made at the time of publication. There are very few standards by which these robots do not seem to be persons, except perhaps in their lack of free will.

The robots lack of free will is their biggest obstacle to recognition as full human beings, and based on the way in which Asimov constructed his narrative, it also requires an examination of what is the appropriate relationship between human and non-human persons. Part of the reasons that Asimov's robots seem so moral, so compassionate and warmer than many of the humans in the story is because their free will is strictly limited by the Three Laws. Since Asimov defines these robots primarily as a technology, not as persons, he built a strict hierarchy into the Laws. The highest priority for any robot is not their own self-interest, or any other priority that they can choose, it is always and inherently the preservation and protection of either humans individually or humanity as a whole. Susan Calvin, as the narrator, seems to agree wholeheartedly with Asimov that this is the appropriate moral order for these robots, and she reacts very strongly in "Little Lost Robot" when the First Law is weakened. She even bases her

claim that robots are the superior beings off of this hierarchical and subservient programmed morality. There are many logical reasons for these Laws, but they are difficult to reconcile with the conception of robots as full persons.

Part of Herzfeld's objection to forming a full relationship with artificial creations, is that they are not humans created in God's image, and therefore humans cannot properly enter into an I-Thou relationship with these beings. She concedes that there is some need to have a relationship, but demands that it must be different from the relationship between two organic human beings. The issue with Herzfeld's argument is that it can be possible for created beings, in fiction, if not in reality, to become full people, and Asimov's robots seem to approach it. However, it is impossible to enter into a fully realized I-Thou relationship with beings who are in an imposed condition of servitude and slavery. The robots are humanity's nursemaids, both in the literal sense with Robbie, and in a more metaphorical sense, with the Machines in "The Evitable Conflict," who have essentially taken on this role of protector for humanity as a whole. However, in a traditional arrangement, the nursemaid is still a servant, and not a powerful one. There is a degree of power over the child, since the child must obey their keeper, but the nursemaid ultimately must answer to a higher authority in the parents. It is also possible, as is seen in "Robbie" that the child has a degree of authority, since she is the one with the officially recognized status, either as "mistress" or as "human." In these initial stories, Asimov argues that humanity needs the robots to care for them, and should trust them as protectors, but never addresses the fact that they are required to remain servants, and that they are never fully recognized as persons, even when they are in control.

It is telling that when the robots become masters of humanity, they do so in a way that is typically coded as "feminine," which is very much in line with their role as the nursemaids. They

do not take control in a militaristic or overt way, they do so in a way specifically designed to prevent anyone from being aware of their power. They exercise a soft power, one of subtle manipulation of events, and concealing their true motives. It is notable that when Stephen Byerley, a male-coded character, discovers that there is an attempt to hurt the robots, his response is to take overt action, "have the [anti-robot] Society outlawed, every member removed from any responsible post."[313] It is Susan Calvin, a female scientist who, despite being depicted as cold and robotic, also succeeds in a very male-dominated field for decades, explains to him that the Machines are far more subtle and far more effective at resolving the issue than he is. She explains to him what the Machines' solution is, and the tells him that, "it would be harmful to humanity to have the explanation [for the Machines' issues] known."[314] She sees that the Machines are using the limited power that humanity allows them in ways their creators never imagined.

For all his attempts at creating a secular, technological vision of artificial intelligence, even Asimov admitted that when writing science fiction questions of religion inevitably arise. In *I, Robot*, Asimov asks much the same religious questions as Mary Shelley did over one hundred years earlier, but asks them for a new generation with a new and changing relationship to technology. When humanity creates a new being, without the help of Aphrodite or the sacred name of the God of the Hebrew Bible, it requires a consideration of what this being's role is in the order of the world. It throws into question the place of humanity, which has been traditionally seen as a privileged place as the most intelligent and conscious creature on Earth. It asks what the real definition of a person is, and to whom rights should be afforded. If the creature is not

[313] Isaac Asimov, *I, Robot* (New York: Bantam Dell, 1950), 268.
[314] Ibid., 271.

made in the image of God, but instead in the image of humans, it becomes the responsibility of the new human creators to find a place for their beings.

In Genesis, God is able to find the first humans "good," whereas the human attempts at creation in fiction are less successful. Frankenstein sees his creation only as a monster, a fiend to be feared and destroyed. Asimov's robots are likewise feared, but, perhaps more damagingly, even their advocates see them only as technology and property. These fictional narratives of creation serve as a reminder of the precarious and complicated position of humanity as both creations and as creators of new technology that is constantly changing the ways in which humans interact with the world. These artificial intelligences encourage a closer examination of the relationship with the other and the potential for I-Thou relationships, especially with those who are frequently disregarded and seen as unable to fully participate in relationships. Thoughtful stories of human creation, from *Frankenstein* to *I, Robot*, require human readers to consider their own complicated status as creations, creators, and participants in relationships, and to ask what it is to be a person and what it is to create.

REFERENCES

Aldiss, Brian W. *Billion Year Spree*. New York: Doubleday, 1973.

Abbas, Niran B. *Thinking Machines: Discourses of Artificial Intelligence*. Hamburg: Lit, 2006.

"Artificial Intelligence." *Funk & Wagnalls New World Encyclopedia* (2015): 1p. 1.

Asimov, Isaac. *Asimov on Science Fiction* Garden City, N.Y. : Doubleday, 1981; 1st ed, 1981.

Asimov, Isaac. "*Gold: The Final Science Fiction Collection*. New York: HarperPrism, 1995.

Asimov, Isaac and Janet Asimov. *It's been a Good Life*. Amherst, N.Y.: Prometheus Books, 2002.

Asimov, Isaac. *I, Robot*. New York: Bantam Dell, 1950.

Asimov, Isaac. "Social Science Fiction." In *Modern Science Fiction, its Meaning and its Future* edited by Reginald Bretnor and John W. Campbell Jr., 158-196. Chicago : Advent Publishers, 1979; 2d ed, 1979.

Campbell Jr., John W. "The Science of Science Fiction Writing." In *Of Worlds Beyond: The Science of Science Fiction Writing*, edited by Lloyd Arthur Eshbach, 86-96. Reading, Pennsylvania: Fantasy Press, 1947.

Čapek, Karel. *R.U.R*. Translated by Paul Selver and Nigel Playfair. New York: Doubleday, 1923.

Dewey, Joseph. "Romanticism." *Salem Press Encyclopedia* (January 2015): *Research Starters*, EBSCO*host* (accessed April 18, 2015).

Encyclopedia of Religion, s.v. "Fiction: The Western Novel and Religion," by Lindsay Jones, accessed 4 April 2015.

"Godwin, William." *Funk & Wagnalls New World Encyclopedia* (2014): 1p. 1. *Funk & Wagnalls New World Encyclopedia*, EBSCO*host*(accessed April 19, 2015).

Gonzalez, Justo. *A History of Christian Thought, Volume 3*. 2nd ed. Nashville, Tennessee: Abingdon Press, 1987.

Henthorne, Susan. "Paradise Lost by John Milton." *Salem Press Encyclopedia Of Literature* (January 2014): *Research Starters*, EBSCO*host* (accessed April 12, 2015).

Herzfeld, Noreen. *In Our Image: Artificial Intelligence and the Human Spirit*. Minneapolis, MN: Fortress Press, 2002.

Ketterer, David. *Frankenstein's Creation*. Victoria, Canada: English Literary Studies at the University of Victoria, 1979.

Krulik, Ted. "Bounded by Metal." In *The Intersection of Science Fiction and Philosophy: Critical Studies*, edited by Robert E Myers, 121-132. Westport, Conn.: Greenwood Press, 1983.

Le Guin, Ursula K. *The Left Hand of Darkness*. New York: Ace Books, 1969.

Levine, George. "The Ambiguous Heritage of *Frankenstein*." In *The Endurance of Frankenstein*, edited by George Levine and UC Knoepflmacher, 3-30. Berkeley, California: University of California Press, 1979.

McCauley, Lee. "The Frankenstein Complex and Asimov's Three Laws." *Association for the Advancement of Artificial Intelligence*. Published May 2007. https://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-003.pdf

Michie, Elsie B. "Marx's Theories of Alienated Labor." In *Approaches to Teaching Mary Shelley's Frankenstein*, edited by Stephen C. Behrendt 93-98. New York: The Modern Language Association of America, 1990.

Milton, John. *Paradise Lost*. New York: Barnes and Noble, 2004.

Morrison, Lucy, and Staci L. Stone. *A Mary Shelley Encyclopedia*. Westport, Connecticut: Greenwood Press, 2003.

Puncher, Jeff. *Brave New Words: The Oxford Dictionary of Science Fiction*. Oxford: Oxford University Press, 2007.

Reichardt, Jasia. "Artificial Life and the Myth of Frankenstein." In *Frankenstein Creation and Monstrosity*, edited by Stephen Bann, 136-157. London: Reaktion Books, 1994.

Schiefelbein, Michael. "'The Lessons of True Religion': Mary Shelley's Tribute to Catholicism in 'Valperga'." *Religion & Literature*, 1998., 59, *JSTOR Journals*, EBSCO*host* (accessed February 1, 2015).

Shelley, Mary Wollstonecraft. *Frankenstein: or, The Modern Prometheus*. 2nd ed. New York: W.W. Norton, 2012.

Shieber, Stuart M. Introduction to *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, edited by Stuart M. Shieber. Cambridge, Mass.: MIT Press, 2004.

Stableford, Brian M. *The Sociology of Science Fiction*. San Bernardino, Ca. : Borgo Press, 1987.

Suvin, Darko. *Metamorphoses of Science Fiction: On the Poetics and History of a Literary Genre*. New Haven: Yale University Press, 1979.

Turing, Alan M. "Computing Machinery and Intelligence." In *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*, edited by Stuart M. Shieber, 67-95. Cambridge, Mass.: MIT Press, 2004.

Warrick, Patricia. *The Cybernetic Imagination in Science Fiction*. Cambridge, MA: MIT Press, 1980.