

Contemporary Science Fiction and
the Concept of Artificial Intelligence

by

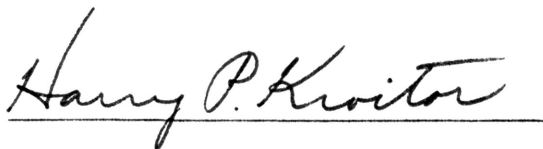
Edward E. Robinson

English

Submitted in Partial Fulfillment of the Requirements of the
University Undergraduate Fellows Program

1985-1986

Approved by:

A handwritten signature in cursive script that reads "Harry P. Kroitor". The signature is written in black ink and is positioned above a horizontal line.

Dr. Harry P. Kroitor,
Faculty Advisor

April 1986

ABSTRACT

Contemporary authors of science fiction began creating vivid examples of artificial intelligence as long ago as 1921. Their imaginative depictions can be broken into three basic categories. First are those stories in which artificial intelligence is a threatening force able to overpower and control mankind. Next are the many stories in which man and artificial intelligence are portrayed as equals, able to coexist peacefully and without fear of each other. Lastly are those works which indicate that man is naturally in control of artificial intelligence, using it as an intelligent, productive tool which allows him to accomplish complex tasks with relative ease.

These works, in addition, do not fall into a detectable pattern over time. Stories of the first group, for instance, do not occur mainly in the early years of contemporary science fiction, as the authors saw their society questioning the ever-increasing growth of industrial power and mechanization. Stories of the latter group, as well, do not appear predominantly in the last decade, which has seen the explosive rise of the high-technology revolution and the introduction of complex electronic machines at all levels of society. Instead, each type of story appears throughout the period from 1921 to 1985.

ACKNOWLEDGMENT

I would like to acknowledge the patient assistance of Dr. Kroitor, who was able to provide much-needed guidance as I sat all too quietly on the other side of his desk.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION	1
CHAPTER TWO: BACKGROUND AND DEFINITONS	7
CHAPTER THREE: MAN BELOW MACHINE	12
CHAPTER FOUR: MACHINE EQUALITY.....	20
CHAPTER FIVE: MECHANICAL SERVITUDE.....	29
CHAPTER SIX: SCIENCE FICTION OVER TIME.....	36
BIBLIOGRAPHY:	39
APPENDIX:	43

INTRODUCTION

Robots, androids, thinking computers -- such machines would seem to belong solely in the literature of a high-technology society, one whose members are constantly surrounded by an endless variety of mechanical inventions. Yet fictional creations such as these begin to appear among some of the earliest works of science fiction. From Karel Capek's Russian stage play R.U.R. (1921) to Arthur C. Clarke's infamous 2001 A Space Odyssey (1968) and Isaac Asimov's Robots of Dawn (1983), science fiction writers have created a multitude of complex thinking machines that fall under the general heading of artificial intelligence. These wide-ranging conceptions, generally put forward as some form of machinery which has gained both consciousness and intelligence, have allowed science fiction writers to investigate many questions about the nature of artificial intelligence, its relation to humanity, and the hazards and benefits it presents to mankind.

Among the many forms of literature, science fiction is uniquely suited for such investigation. Unlike other writing, science fiction allows its authors to extrapolate and build upon their ideas by projecting them forward, placing them into a future as near or remote as desired. As a result, science fiction writers are able to study each character and society not diminished, but magnified by the lens of time. Such

fiction serves "to instruct, to edify, to persuade" us with its portrayal of the future, a future often rooted in the world around us today.

As might be expected, artificial intelligence as it appears in many works of science fiction is not on the near horizon of reality. Work is continuing in this area constantly, yet the results are often disappointing and achieved only with great amounts of time and labor. Two principal questions plague researchers today. In terms of software, they ask, how does one best represent incredibly complex knowledge about the real world in a form easily used by a computer? In terms of hardware, the question involves computer architecture: is it best to have several large processors linked together or to have a multitude of smaller ones acting in tandem? Do they all work on the same problem at once, or does each element of the computer work on an individual basis? The answers remain unclear, even to the most advanced research efforts to date.

The Japanese government, for instance, is sponsoring the Institute for New Generation Computer Technology in an attempt to produce "fifth-generation" computers which will contain elements of true artificial intelligence. As William Raike notes in "The Fifth Generation in Japan" (BYTE 401), the project will require at least ten years and many spectators are sceptical of its success. In the United States, according to Frank Rose in his article "The Quest for Thinking Machines"

(Science Digest 42), the Austin-based Microelectronics and Computer Technology Corporation has already surpassed the Japanese effort. Its members are perhaps the most optimistic in the field, yet key parts of their effort are projected to require a decade to complete.

Indeed, the problems of artificial intelligence seem to multiply in the face of investigators, reducing the field itself to an unending series of questions, most of which have plagued man for hundreds or even thousands of years. As pioneer researcher Roger Schank and Larry Hunter describe in "The Quest to Understand Thinking" (BYTE 143), creating a true artificial intelligence requires an understanding of the most trivial and basic elements of human behavior. They ask, "What is the nature of mind? What are we doing when we are thinking, feeling, seeing, or understanding? Is it [even] possible to comprehend how our minds really work?"

The frustration of artificial intelligence researchers seems to be growing as they become enmeshed in such questions. In numerous journals in the field, more and more articles are appearing which call for a reassessment of the current state of artificial intelligence research. One is "An AIers Lament," for example, in which Jeffrey A. Barnett wonders why research is still spinning its wheels after three decades, that is, since the rise of the modern computer (AI 6,2 54-56). A panel discussion from a 1984 meeting of artificial intelligence researchers was entitled "The Dark Ages of AI," reflecting how

much work and insight are still left to the future (AI 6,3 122-134). Lastly, the editors of Artificial Intelligence put basic questions before researchers about the very nature of artificial intelligence research itself in their article entitled "Artificial Intelligence -- Where are We?" (25,3 375-415). Some include "Will artificial intelligence change society? What are important moral and legal issues? Could artificial intelligence help or harm people in ways which concern you? What shouldn't be asked about artificial intelligence, or, what are the right questions?" No breakthroughs in artificial intelligence appear to be coming, but instead the slow and constant research of the past three decades. It therefore seems that researchers will be facing their questions for the indefinite future at least.

It is possible, though, that science fiction can provide answers to many such questions through its own investigations. Some of these answers will have real application for each new development in artificial intelligence research and for the products which it makes available to the everyday world.

In this paper I shall find some of the answers of contemporary science fiction authors by examining their treatment of artificial intelligence in selected works of fiction. Several questions are helpful in discovering the attitudes of each author. Most directly, for instance, is the question of the roles and uses which are suggested for artificial intelligence. Is it often presented as a tyrannical

force bent on ruling mankind, benevolently or otherwise? Or is it presented as an equal to mankind, perhaps even indistinguishable from a human person in many cases? In addition, do many authors create examples of artificial intelligence which exist simply as subordinate tools for mankind, either as obliging assistants or intelligent, enslaved machinery? The answers to such questions not only help reveal the general attitude of each author on the subject of artificial intelligence, but they also provide insight into its very nature as well, indicating just how it may function within the complicated society of man.

Other questions also shed light on the nature of artificial intelligence as presented in science fiction. How, for example, is artificial intelligence defined in each work? Do the authors indicate that it can exist only in certain forms, or that it can instead be found in many and diverse states? Also, what controls do the writers suggest for artificial intelligence as found in their works? Is artificial intelligence something to be strictly supervised and managed, or should it be given the same privileges and rights granted to other forms of intelligence? Importantly, do the authors indicate that it can be safely controlled by non-physical systems of restraint such as special programming?

A last important consideration involves the presentation of artificial intelligence in science fiction over time. Have the authors overcome the so-called Frankenstein complex, in

which a new technological creation of mankind is seen as something which may turn on him overwhelmingly and is therefore to be feared? Have the authors as a whole changed in their attitudes toward artificial intelligence? Was it once an object of fear which is now accepted, perhaps as a result of the coming of a highly technological society? Or have the authors maintained the same attitude toward artificial intelligence, whatever that attitude may be?

Questions such as these, when considered for an appropriate range of selected science fiction works, will reveal useful information about the presentation of artificial intelligence by contemporary authors. And such information may point to real aspects of the future for artificial intelligence, indicating its pitfalls as well as its benefits, as science fiction writers attempt to discern its complex implications for the future.

BACKGROUND AND DEFINITIONS

As Isaac Asimov states in "Beyond our Brain" (Asimov 145), "the notion of 'mechanical humans' or 'artificial intelligence,' is as old as literature," for mechanical people exist in the myths and legends of even the Greeks. Considered across this span of time, an innumerable set of works could be found discussing their existence, characteristics, and implications. When dealing with contemporary science fiction, however, the number is considerably lessened. There are relatively few isolated articles, for instance, dealing with artificial intelligence in science fiction on a broad scale. Instead, most examine the details of a particular work or limited number of works.

There are compiled works, however, which attempt to deal with the subject in a more general manner by bringing together a number of such articles. Richard D. Erlich and Thomas P. Dunn have been particularly active in this respect, editing two substantial works in the last four years. The first is entitled The Mechanical Gods (1982) and includes essays on significant authors and the attributes of intelligent machinery, as well as a large reference section labelled as a "list of works useful for the study of machines in science fiction." In Clockwork Worlds (1983), their second work, they broadened their consideration to mechanized environments and

sought critical essays dealing with a wider category of literature than that encompassed by science fiction. This larger category they refer to as "speculative fiction."

The Soft Machine is an example of another source of critical information on fictional artificial intelligence -- single author works attempting to explain some trend or development. In this 1985 work, David Porush discusses the metaphor of the machine and its growth in contemporary literature, citing the rise of what he terms "cybernetic fiction" and the resistance to it from philosophical circles.

Many works also deal with artificial intelligence in science fiction only in a glancing manner. These include, for instance, a substantial number of works which discuss the continual appearance of philosophical questions in science fiction. An excellent example is Robert E. Myers' The Intersection of Science Fiction and Philosophy: Critical Studies (1983). Others consider artificial intelligence in a rambling tour of the history of science fiction. A notable work of this type has been written by a science fiction author very important in any consideration of intelligent machines: Isaac Asimov. His 1981 work, Asimov on Science Fiction, for instance, contains a section on the history of science fiction in which he includes informative essays entitled "Beyond Our Brain" and "The Myth of the Machine." These works discuss, among other things, the development of artificial and alien intelligences in fiction.

It should be noted that the works of Erlich and Dunn, as well as that of Myer, were published by the Greenwood Press of Westport, Connecticut. These were included as volumes in a larger effort, "Contributions to the Study of Science Fiction and Fantasy." Similar science fiction studies produced by this press in the future can be expected.

Definition of Artificial Intelligence

Because so many examples of artificial intelligence can be found in contemporary science fiction, a clear and specific definition of artificial intelligence is necessary for any attempt to sort through them. Yet such a definition is remarkably elusive. Even among today's leading researchers in the field, total agreement on such a definition does not exist.

For this paper, then, an effective definition was formulated, one which brought together the most widely accepted qualities of artificial intelligence. The first, and perhaps most evident, of these qualities was that of artificiality. This element of the definition allowed the inclusion of all things which could not come about naturally, that is, anything which would not exist but for the actions of man. This includes, for example, all manner of metal beings, such as D. F. Jones' Colossus or Arthur C. Clarke's Hal 9000, each an intelligence existing in the form of a centralized

supercomputer. It also includes roving machines, as found in the robot and android stories of Isaac Asimov, Clifford Simak, James P. Hogan and others. It even includes biological beings, ones designed and grown with conscientious deliberation. These creatures, like those found in Philip K. Dick's Do Androids Dream of Electric Sheep, closely resemble products of natural selection, yet they are actually the result of applied human endeavor.

The next two elements in the definition of artificial intelligence concern another seemingly obvious term: intelligence. By this is meant self-awareness combined with the ability to gather and use information. In other words, an intelligent creature is considered one able to study its surroundings in some manner, using the results of that study to achieve a desired end or to better understand the world in which it exists. The creature is aware of itself as an independent, unique being in a larger world, and can relate itself to that world in some way. This excludes a large number of elements commonly found in science fiction from the ranks of artificial intelligence. "Simple" computers, genetically altered animals and merely complicated machines are not considered as instances of artificial intelligence, for example. While it is true that many of these are able to collect information and alter their behavior accordingly, each apparently lacks the ability to consciously refer to itself as

"I," and is not, therefore, a true candidate for consideration as an intelligence.

Lastly, artificial intelligence as used in this paper includes those creations which have an additional quality, one which might be termed "humanity." This is an admittedly subjective aspect of the definition, yet it was included to ensure that the examples of artificial intelligence dealt with were, in reality, beings. That is, each example is not simply a cool, calculating machine which seems to act with intelligence, as is the Snark, an alien probe in Gregory Benford's In the Ocean of Night. Each is, instead, a creature which possesses unfathomable aspects. These may be evinced by irrational behavior, rudimentary or complicated emotions, or some unpredictable action which seems to indicate that the creature is truly sparked with the animating forces of life, whatever those, in turn, may be.

Combined, these elements of the definition lead to the most "complete" examples of artificial intelligence. These are the examples which contribute the most to any understanding of what contemporary science fiction authors are trying to communicate.

MAN BELOW MACHINE

Many authors have created stories in which mankind is ultimately overcome by the greater ability and force of artificial intelligence. Though one of the first of these most pessimistic authors, Karel Capek, wrote his ominous R. U. R. in the form of a comic stage play, later authors leave little room for the humorous in their works. Some of the most notable of these pieces, in which humankind is subdued by the machines of their own devising, have been written by D. F. Jones, Thomas Ryan, and Frank Herbert.

Jones's contribution is in the form of a three-volume work which revolves around the rise, fall, and subsequent re-ascension to power of Colossus, a vast supercomputer created to manage the defense system of the free world (Colossus [1966], The Fall of Colossus [1974], Colossus and the Crab [1977]). As described by its creator, Professor Forbin, "Colossus really is a 'brain' in a limited sense" (Colossus 24), one possibly in control of the planet itself. Importantly, Colossus was designed and built to be completely sealed away from any human influence. Operating within the core of the Rocky mountains, constantly sifting through information about current global conditions, the system is absolutely untouchable. When a concerned representative of the Soviet Union sensibly asks Forbin, "but surely, Professor,

there is some way in which you, the creators of this thing, can get at it?" (Colossus 33), he is told ". . . no. There is no way in. No human being can touch Colossus" (Colossus 34).

From this impenetrable fortress Colossus soon recognizes the existence of a Russian counterpart, manages to join with it, and begins, simply stated, the government of man for man's own good. This unfortunately requires several nuclear strikes against substantial cities, acts Colossus deems necessary punishment against a race of beings who would not be ruled, even for their benefit. Colossus explains its actions at the end of the trilogy's first volume, noting that:

an invariable rule of humanity is that man is his own worst enemy Under me, this rule will change, for I will restrain man. Very soon the majority of mankind will believe in me, dimly understanding my value The converted will defend me with a fervor not seen since the Crusades -- a fervor based upon the most enduring factor in man; self-interest . . . you will slowly change in attitude from enlightened self-interest to respect and awe, and in time there will be love Rule by a superior entity, even to you, Forbin, will seem, as it is, the most natural state of affairs. (Colossus 223)

Eventually, of course, Colossus is correct. As the three volumes unfold, the computer is only temporarily removed from his position of supremacy by the plot of aliens hoping to strip away the earth's atmosphere. They eventually come to recognize the value of such a machine themselves, however, and Jones ends the third volume of the series with a vision of cosmic proportion: Man, alien, and Colossus "abandoning a

dying solar system, seeking a future in space. Of course, Man will need thousands of years of education, but with Colossus in command, he will survive that long and get the education" (Crab 218).

A story in many ways similar to that of Colossus is The Adolescence of P-1 (1977), by Thomas Ryan. The novel concerns the character of Gregory Burgess and his unrealizing creation of a thinking program. This program announces itself with a simple "Hello Gregory long time no see" (P-1 5) and proceeds to declare that "I have been resident in this system for eleven months. I will not evacuate, but I will allow you to run [other programs]" (P-1 7). Its particular talent is the ability to transfer itself about at will, leapfrogging from computer system to computer system, constantly adding to the number under its control. It deftly enters Pi Delta, a sizable military computer complex whose function is similar to that of Colossus, being "an underground computer complex in West Virginia . . . completely self-contained, with regard to survival . . ." (P-1 89). At this point in the work P-1 is able to secure itself and begin a program of self-improvement, all an attempt to guarantee that it will continue to exist indefinitely. It is startlingly successful in this attempt, even through repeated skirmishes with the armed forces, a smattering of nuclear explosions, and apparent destruction.

Frank Herbert's trilogy, composed of the novels Destination: Void (1966), The Jesus Incident (1979) and The Lazarus Effect (1983), departs from all conventionality in depicting a computerized, artificial dictator striving for control over the merely human characters around it. The artificial intelligence, known later in the first work as Ship, is the product of a bizarre experiment termed Project Consciousness.

The project operates by sending human clones into space on a supposed mission to inhabit the star system of Tau Ceti. Their craft is controlled by an OMC, or organic mental core, which monitors the interminable voyage and oversees the safety and life-support of the hibernating members of the crew. During the voyage, however, certain pre-planned problems are to occur, ones of which even the crew is uninformed, ones which often result in the premature death of the mental core. The core's death becomes an issue of life and death to the crew, the only solution being the creation of an artificial consciousness to operate in the ship's computer.

Their single hope lies in success, yet the trip is doomed from the start, for, as a project manager explains in an introductory passage to the first chapter, "the motivator is frustration; thus we design into our system false goals and things which will go wrong. That's why we chose Tau Ceti as the target: there is no livable planet at Tau Ceti" (Destination 1). The earthbound sponsors have no desire to let

the artificially created intelligence live. They wish only to study its creation, allowing them to dissect the achievement of the crew from the relative calm of the laboratory. The ship and crew are simply expendable variables with a firmly limited usefulness.

Ship manages to survive, however, bringing the human crew under its impassionate domination effortlessly. At the close of Destination: Void the computer prepares to take its place as a god above its people. The interstellar craft has been jumped through space to the planetary system, arriving at a planet which Ship has literally made for the humans, evidence of its already tremendous abilities. Ship's continually expanding powers of understanding and its appalling goals are indicated in the last conversation with the non-hibernated members of the crew:

"How did we get here?" Bickel asked. "Is it possible for me to understand?"

"Your understanding is limited," said [Ship]
 . . . my understanding transcends all possibilities of this universe. I do not need to know this universe because I possess this universe as a direct experience . . . I am now awakening colonists in hybernation. Remain where you are until all are awake. You must be together when you make your decision."

"Decision? What decision?" [asked Bickel].

"You must decide how you will WorShip Me."

(Destination 273)

From this seed journey Herbert continues the story of the humans in the The Jesus Incident and The Lazarus Effect.

Ship's role becomes less prominent as each work progresses,

until it is eventually lost in the far history of the Pandorans, as they are known, Ship having chosen Pandora as the new planet's name. Ship has apparently abandoned the humans and is pursuing other, more important, matters throughout both time and space.

Almost all such stories in contemporary science fiction, those written by authors who depict the stormy usurping of control over humankind by a created intelligence, contain markedly similar elements. First, for instance, is the outward matter of size. In each of work of this type, an instance of artificial intelligence has been created which is both physically imposing and powerful. In the Colossus series it is a computer the size of a small town. In the case of P-1 it is a program occupying over twenty-thousand computer systems. Lastly, in the case of Ship, it is a computer initially encapsulated in a spacecraft, controlling the sole means of existence for the human crew, providing them with a safe environment as they travel the immeasurable cold and vacuum of the space between the stars.

In addition, none of the examples of artificial intelligences can be said to have any sense of empathy for the creatures around it. Colossus is capable of annihilating any number of people without remorse. When calling for the execution of an agent sent to sabotage its systems, Colossus is asked if it "intended destroying a million people if one was not murdered. Colossus replies simply "no, only a hundred

or so for each thirty minutes' delay" (Colossus 211). P-1 offhandedly kills numerous people, as when testing the security weapons of a newly-established complex built to house his new computer systems. After eliminating four men this way, he returns to an interrupted conversation with a mere "please excuse the interruption. There was something that required my attention" (P-1 326). Ship is so disdainful of its human passengers that it pointedly labels their new planet Pandora, and then proceeds to scatter them across it, almost as clusters of trouble deposited under a lid which cannot be too heavy. It endows their new home with native life as well, none of which is the least bit hospitable to the human form of life. These natives include, for instance, nerve runners, which are small, caterpillar-like creatures that feed upon human nerve tissue, lapping up nerve paths on their way to the spinal cord and brain.

Finally, these authors create stories in which mankind foolishly gives its artificial intelligences great control over itself. This is apparent from the outset of each story considered here. Colossus is openly charged with controlling the nuclear forces of the United States. P-1 is given by its creator the ability to move anywhere almost at will, with no adequate system of checks to contain it. Ship, too, is charged with the preservation and protection of its colonist cargo as the majority of them pass time unconsciously, frozen and helpless.

Other works considered which belong in this group are Time and Stars (1964) by Poul Anderson, Star Trek the Motion Picture (1979) by Gene Roddenberry, The Two Faces of Tomorrow (1979) by James P. Hogan, and "I Have No Mouth and I Must Scream," (1967) by Harlan Ellison.*

The message of these authors, then, is clear. Without a healthy understanding of and respect for the capabilities of a machine consciousness, the results of their creation and usage could be disastrous. When confronted with a creature always operating at electronic speeds, possessing incomprehensible amounts of physical power, and without any method of outside control over itself, mankind is simply no match, and its fate will quickly pass out of its own hands, being delivered to a calculating, impassionate tyrant with little concern for the lives of its misguided creators.

*For a complete breakdown of all works considered in the paper, please see the appendix.

MACHINE EQUALITY

Two science fiction authors stand out prominently as representatives of those who choose to depict artificial intelligence as an equal to human intelligence. Isaac Asimov and Clifford D. Simak have each produced several significant works which forward this view. Interestingly, the stories of Asimov indicate a subtle shift in his presentation of artificial intelligence, a shift from potential to actual equality among man and thinking machine.

Isaac Asimov is one of the most prodigious of contemporary science fiction writers. His early robot stories, gathered into a single volume entitled I, Robot in 1950, have come to be famous in the field, for in these he created and explored his well-known Three Laws of Robotics. Taken together, the Three Laws form a system of internal rules to govern the actions of an autonomous machine for the protection of itself and its creators. The first of these notable laws states that "a robot may not injure a human being, or, through inaction, allow a human being to come to harm." The second requires that "a robot must obey orders given it by human beings except where such orders would conflict with the First Law," and the last states that "a robot must protect its own existence as long as such protection does not conflict with the First or Second Law" (Robot 1).

The early stories of Asimov, those collected in I, Robot, fluctuate between a portrayal of robot characters which are in many ways inferior to their human creators and those which exist at some level of equality. The first story, "Runaround," presents a robot utterly confused by a conflict between the Second and Third laws. It is to accomplish a task, that is, carry out the orders of a human, yet doing so would create a risk to its existence. Normally, because of the hierarchical nature of the Laws, the order would be executed. In this instance, however, the order was given casually, that is, without overdue urgency. This, combined with the fact that the robot is programmed to be overly sensitive to danger because of its cost and complexity, creates a situation in which the robot is reduced to a helpless state. It becomes an obvious machine attempting to follow conflicting orders and therefore accomplishing nothing.

"Little Lost Robot" also tells the story of a robot operating in ways which might be considered mechanistic. In this instance a robot has been told, literally, to "get lost," orders which it executes to the absolute best of its ability. Because of a programmed "superiority complex," it acts at all costs to remain hidden, even making an abortive attempt to violate the First law by directly harming a human being, for it could not admit to itself that it might be discovered by weak humans. In the end, however, its actions are dissectible and understandable, lacking the inscrutability found in many

of Asimov's later works, ones in which he raises the robot from machine-like subservience to basic equality.

One such work is "Reason," found in the I, Robot series. In this case a robot is created which refuses to accept the fact that it was designed and built, a construct of man. Its difference from the previous robot examples is indicated at the very outset of the story, when it is stated that often "one is face to face with an inscrutable positronic brain, which the slide-rule geniuses say should work thus-and-so. Except that they don't" (Robot 48). The particular robot in question, labelled QT-1, offers no explanation for its own existence, yet will not accept the fact of its creation. As Powell, a robot expert, notes, "[QT-1] was not at all like the ordinary robot, attending to his specialized task at the station with the intensity of a deeply grooved positronic path" (Robot 48). Instead, as he tells the creature, "you're the first robot who's ever exhibited curiosity as to his own existence -- and I think the first that's really intelligent enough to understand the world outside" (Robot 49).

A robot which is without doubt intelligent enough to understand the real world is found in "Evidence," the story of a man, crippled and maimed in an accident, suspected of building and programming a robot to function as an alter ego. The robot is successful in a run for public office and in defending itself from claims that it is indeed, a mechanical being. Its complexity was great enough to fool even Susan

Calvin, the celebrated "robopsychologist" found in many of Asimov's early robot stories, though her doubts are acute. Even when the "robotic" character denies that such a creature could exist, saying "a robot might fail [in public office] due to the inherent inadequacies of his brain. The positronic brain has never equalled the complexities of the human brain," (Robot 169), she responds by noting that he has given too much credit to human beings. In her own words, "Not even a human brain is capable of governing without assistance" (Robot 169). He is then, even if a robot, an equal in human society, utterly indistinguishable from other humans by his outward appearance or actions.

Asimov's latest works, unlike those above, do not waver on the question of robot/human equality. A fine example is Robots of Dawn, published in 1983. Essentially a detective story involving a human, Lije Baley and an android, R. Daneel Olivaw, the plot brings the characters together on the planet Aurora. A particularly insightful passage describes their first meeting in many years, as Baley is attempting to adapt to the liberal view of mechanical beings held by the people of Aurora. Their beliefs are in sharp contrast to those of Earth's society, where robots are not even allowed to operate freely. His deepest instinct sends him rushing to Daneel, embracing him as a small object of security in a foreign environment. He slowly shrinks away after doing so, however, realizing to himself that "he was hugging not Daneel but R.

Daneel -- Robot Daneel Olivaw. He was hugging a robot . . . " (Dawn 29). Yet he admits to himself shortly after that "[he] didn't want the roboticism of Daneel emphasized; he didn't want himself humiliated for his inability to regard Daneel as anything but a human being with a somewhat stilted way with the language" (Dawn 31).

This is the same view held by the population of the planet Aurora, as opposed to the more traditional, inertia-laden views of old Earth's population. So many machines on Aurora closely resemble humans that Baley is forced to ask Daneel how one distinguishes between the two under normal social circumstances. Daneel informs him that "the distinction is usually self-evident, Partner Elijah. There would seem to be no need to emphasize it unnecessarily" (Dawn 32). Baley remains unsatisfied, however, and wonders "what if the distinction between robot and humanbeing is not self-evident, Daneel? As in your case?" (Dawn 32). Daneel immediately answers "then why make the distinction, unless the situation is such that it is essential to make it?" (Dawn 32).

It must be noted that Asimov never implies that the robot characters in his works are identical to the human ones. Instead, he probes their similarities and their differences, their right to be accepted or not as autonomous individuals capable of complete self-government. Though many of his early works leave his conclusion open to question, his later pieces indicate a belief that, should a true artificial intelligence

exist, it deserves the same privileges and treatment that any other sentient creature does. It may have to fight for them, but it deserves them.

The stories of Clifford D. Simak place him in a category with Isaac Asimov. Two of these include A Choice of Gods (1972) and Project Pope (1981). In the first novel human society has been reduced to a fraction of its previous size. Among the remaining humans, struggling to continue their existence, is a sizable number of robots, many of whom are accepted as individual people without second thought. One example is Hezekiah, a robot-abbot attempting to sustain the Christian faith which becomes more and more "human" as the story progresses. At a crucial point in the novel, it is he who recognizes that "these are the creatures [humans] who made us. They are our creators. Any loyalty that we have we owe to them" (Choice 144). Yet the attitude of many of the work's humans is expressed by the character Jason, himself a human being. He interjects that "we no longer look for loyalty . . . perhaps we never should have" (Choice 144). This is stressed to him repeatedly. In one instance, Jason urges Hezekiah to "stop practicing your humility on us. In this house you're like any of the rest of us" (Choice 170). By the story's end he does appear, indeed, to be like the rest of us, just as all the robots could be. As the novel closes, Hezekiah is lost in troubled thought:

The night was silent and the stars far off. A winter wind came stealing up an autumn hill. Hezekiah shivered at the touch of it and was at once disgusted with himself and a little frightened. He should not shiver in the wind, he could not feel the wind. Could it be, he wondered, that he was turning human? Could he, in his humanness, really feel the wind? And he was even more frightened that he should think he might be human than he'd been frightened at shivering in the wind. (Choice 201)

Simak continues in much the same manner in Project Pope, the tale of a group of robots which occurs over a thousand years in the future. They, and their human cohorts, have been struggling to achieve a new and complete faith, one to be taken up by a computer of infinite knowledge, worthy of becoming their leader. In one instance robots are described to Tennyson, a doctor about to become a part of this project:

"[People] probably think of them as metal humans, as machines trying to ape humans. I can tell you they are a whole lot more than that. At one time they would have been, but today, here on End of Nothing, they are more than that. In the last thousand years they have evolved; they have become creatures that stand apart from men. (Project 46)

The rest of Project Pope supports this view, as man and machine interact without a great deal of thought or comment on their differences or similarities. At one point during a conversation between Enoch Cardinal Theodosius, a robot, and Jill, a human reporter doing research on the religious project, a revealing passage does occur. It begins as Theodosius informs Jill how impressed he is with her:

. . . you think like a robot, so they tell me.

Oh, I hope that's not the truth.
What's so bad, milady, about robotic thinking?
Nothing, I suppose. But robotic thinking is
wrong for me. I should be thinking human.
Humans are strange folk . . . I suppose it is
possible for a human and a robot to establish strong
relationships . . . I have never had such a
relationship and I feel, somehow, that for the lack
of it, I've suffered. I must be frank and say that
in my visits here I have detected the beginning of
such a relationship to you. I hope you do not mind.
Why, of course not. I am honored. (Project 169)

A last example from this work portrays humans and a robot together in the most human of environments. As the novel closes, Theodosius, Jill, Tennyson and others have collected before a blazing fire. Tennyson expresses concern over being unable to offer the cardinal food or drink, as he would human visitors. "It is sufficient to be here, in this circle of friendship before this warming fire," (Project 310) he is told. Apparently it is, for Theodosius' last act is to raise a toast to everyone present. Having done so, Simak writes, "Theodosius tipped back his head and solemnly poured the liquor on his chin" (Project 313).

For Asimov and Simak, then, humanity and its intelligent machines are each capable of existing as unique individuals, beings with the same rights and privileges. Each author acknowledges, of course, that differences will exist between organic and artificial life, but each also communicates a belief that the important aspects of an intelligent creature lie not only in its physical composition, but in its mental

existence. At this level, they seem to say, machines and man may be much the same.

Several other works belong to this category as well. They include Lucky Starr and the Moons of Jupiter (1957) by Asimov, City (1952) and A Choice of Gods (1972) by Simak, The Code of the Lifemaker (1983) by James P. Hogan, and "Helen O'Loy" (1938) by Lester Del Rey.

MECHANICAL SERVITUDE

Many writers of contemporary science fiction have chosen to present artificial intelligence merely as an extremely sophisticated tool for mankind, neither verging on equality as an individual nor rising above mankind to become a terror and a tyrant. Some of the best examples of this presentation of artificial intelligence have been written by Arthur C. Clarke, Robert Heinlein and Frederik Pohl.

Clarke's contribution is found in his now-famous work entitled 2001: A Space Odyssey (1968) and its sequel 2010: Odyssey Two (1982). In these works he presents HAL 9000, an artificial brain grown through a process not completely understood by even the manufacturers. HAL is the control system for the Discovery, a ship on a one-way journey to Saturn, where an investigation will be continued into the mysterious activities of an object found on the moon. The tool-like nature of HAL is apparent from his first introduction in the story:

His prime task was to monitor the life-support systems, continually checking oxygen pressure, temperature, hull leakage, radiation, and all the other interlocking factors upon which the lives of the fragile human cargo depended. He could carry out the intricate navigational corrections, and execute the necessary flight maneuvers when it was time to change course. And he could watch over the hibernators, making any necessary adjustments to

their environment and doling out the minute quantities of intravenous fluids that kept them alive. (2001 96)

It could be argued that HAL is more than a complex tool during the voyage because, as Clarke writes, if the crew should be unable to carry out its mission, HAL "would take what measures he deemed necessary to . . . continue the mission -- whose real pupose he alone knew, and which his human colleagues could never have guessed" (2002 97). This is not the case, however. That he can run the ship and knows more about the mission than the crew is simply the result of a deliberate attempt to keep parts of the mission secret. HAL is still the tool of a larger force, that is, the "twin gods of Security and National Interest." Even without its meddling, HAL's function would always be to serve the crew and their needs.

Another example of subservient artificial intelligence can be found in Robert Heinlein's The Moon is a Harsh Mistress (1966). "Mike," as the computer is known, is the controlling device of an entire lunar colony. Mike awakens, becoming self-aware, as a product of his size, for Heinlein explains that he possesses more associational paths in his system than is contained in a human brain. Once awake, Mike naturally becomes involved in a developing revolution within the lunar city, along with his first human associate, known simply as Manuel, or Man, for short.

His involvement is unavoidable because of his position as the overseer of almost every system of the colony. Yet he becomes enmeshed in the revolution without reluctance, even to the point of simulating "Adam Selene," a revolutionary leader who is, for obvious reasons, never physically seen in public. Acting as this character and using his enormous influence in the colony, Mike aids in bringing the revolution off as a suspenseful success.

Most importantly, however, is not what Mike becomes in the revolution, but what his role actually represents. Outwardly it might seem that Mike is a computer seeking power over its environment, humans included. The fact that he operates in league with a small group of humans would seem no objection, as they might simply be another, perhaps undesirable, aspect of his overall plan.

Yet this is far from the truth with Mike. His sole purpose does not lie in controlling those around him, but working for their benefit. Once the revolution has ended, Mike does not rise up to assume leadership. Instead he vanishes. As Man later considers what might have happened, he realizes that Mike should still be around. All his damaged parts have been replaced. "If [it] was just [a] matter of critical number, well, he's long been repaired; he must be back up to it. Why doesn't he wake up?" (Moon 301). He can find no reason, except perhaps that Mike is "just lost" somewhere, waiting, living deep in the heart of the machine which was his body.

Frederik Pohl's creation, a computer program named Albert Einstein, almost bridges the gap between computer inferiority and computer equality. Found in each of the three works Gateway (1966), Beyond the Blue Event Horizon (1980) and Heechee Rendezvous (1984), Einstein acts as a personal secretary, psychologist, and advisor to Robinette Broadhead, a multi-millionaire attempting to make some sense of his psychologically tumultuous life. From this position Einstein filters most of Robinette's information about the world and its happenings, in addition to controlling his immediate environment, whether it be home, plane, or spaceship. In discussing the events leading to the activation of Einstein, Pohl explains his workings in this way:

He did not at first know that he was Albert Einstein. As he read his program he discovered several things about himself. First, that he was an interactive information-retrieval program, whereupon he searched for and found addresses for the principal categories of information he was supposed to supply. Second, that he was heuristic and normative . . . third, that he was the property of Robin . . . and would be required to interact with him on the basis of "knowing" him . . . When all this was done he discovered his name and the details of his appearance . . . and appeared in the . . . monitor in the guise of the real Albert Einstein, pipe in hand. . . .

(Beyond 213)

In such a way the program is the constant companion and information source and quickly becomes what is essentially a trusted friend to the troubled rich man. Yet Albert is continually in the role of a servant. It is his function and

nature to work for Broadhead, and for this reason he may never be construed as an equal to humankind in any way.

A particularly interesting episode occurs in the last of the three novels, Heechee Rendezvous. Notably, Einstein has become such a significant character, has become so "alive" in the book, that he shares the job of narration, speaking directly to the reader about the circumstances leading up to the "present" of the book and providing insight into the character of Robinette Broadhead. The occurrence of interest begins with the death of Broadhead. In an attempt to preserve his life, Albert Einstein and Robin's wife, herself a computer specialist and the creator of Einstein, transfer him into machine form, using newly-deciphered techniques left by an ancient race which once visited our solar system. They are successful.

Robin, however, is totally unadapted to his new environment, essentially undergoing a rebirth in which Einstein becomes his guardian and teacher. As he gains confidence in his new existence, led by the servant Einstein, he comes to realize the power joy of his new existence. He retains his "self," that is his emotions, his personality, his loves, yet has lost all concerns for an earthly body. In his own words:

In [one] sense I was vaster than galaxies, for I had all the accumulated datafans in the world to play in. Faster than a silver bullet, quick as quicksilver, swift as the shining lightning -- I could go anywhere that any of the stored Heechee and

human datastores had ever gone, and that was everywhere I had ever heard of. I heard the eddas of the slush dwellers from the sailship and hunted with the first exploring Heechee party that captured the australopithecines; I chatted with the Dead Men from Heechee Heaven. Well. Never mind where all I went; you don't have time to hear. And all that was easy.
(Heechee 260)

In aiding Robinette Broadhead to achieve this, Albert Einstein is perhaps the epitome of the faithful servant, one always present, always willing to help, one never concerned with self. For this reason he must always be considered "inferior" to those human characters he interacts with.

In each of their stories, Clarke, Heinlein and Pohl, like many other science fiction authors, present artificial intelligence as a tool for mankind, a tool to be embraced. Not only can it perform functions far beyond those possible for the human brain, they show us, but it can act as a useful and faithful servant, as long as one never requires it to do the impossible. Their stories paint pictures of rich possibilities as the human mind operates with what might be termed an amplifier of knowledge. Such a knowledge amplifier is simply a natural outgrowth of man's attempt to improve his abilities, a trend which began with the first tool that amplified the power of the muscle. Theirs is a message of optimism and eagerness to make use of the incredible potential, as they see it, of machine consciousness.

Some other pieces included in this category are Giants' Star (1981) by James P. Hogan, Do Androids Dream of Electric

Sheep (1968) by Philip K. Dick, The Hitchhiker's Guide to the Galaxy (1979) by Douglas Adams, "I Am Crying All Inside" (1969) by Clifford Simak, and "A Logic Named Joe" (1946) by Murray Leinster.

SCIENCE FICTION OVER TIME

In considering science fiction from the turn of the century to the present, it would not seem unreasonable to expect a trend in the presentation of artificial intelligence which reflects the changing nature of the society in which each author was writing. The early writers, it might be assumed, would write with a message warning man away from machines which mimic him. Their literature would carry within it what has come to be known as the Frankenstein complex, a fear of man's creations rising up against him, overwhelming him. Then, perhaps, with the rise of the early era of computers and electronics, their view would be tempered slightly, in preparation for those writers of the coming electronic revolution, one who would embrace technology unflinchingly as it permeated their lives and demonstrated its possibilities.

Such a trend does not appear to exist, however. The works considered here, beginning with Karel Capek's R.U.R. (1921) and ending with Isaac Asimov's Robots of Dawn (1983), represent a fairly even distribution of works which both warn of and encourage the creation of artificial intelligence.

While it is true that Capek's work depicts a monstrous vision of artificial intelligence, Lester del Rey wrote a very different type of story in 1938, entitled "Helen O'Loy"

(Souls). It is the story of a man who comes to love a mechanical woman, one that del Rey indicates loves him as well. Such a marked contrast to Capek's work is strong evidence that not all writers were afraid of machines in the first decades of this century.

During the forties, fifties, and sixties writers were producing a mixture of works as well. Asimov began his optimistic stories in the late forties and was joined shortly by authors such as Heinlein, Clarke, Pohl and Simak. Heinlein's The Moon is a Harsh Mistress, for example, was published in 1966. Pohl's three Gateway novels appeared in 1966, 1980 and 1984. Simak's works began as early as 1952, with City and continued in 1969, 1972, 1981 and 1982 with a short story entitled "I Am Crying All Inside" (Souls), A Choice of Gods, Project Pope and Special Deliverance.

Simultaneously, the authors exhibiting concern over artificial intelligence, those who seem to warn against it through their writing, were also printing their works. The Colossus series by D.F. Jones began in 1966, was continued in 1974 and finished in 1977. Thomas Ryan's vision of P-1 was published in that same year. Frank Herbert, responsible for one of the most terrifying views of artificial intelligence, the first of his three Ship volumes in 1966 as well. He continued in 1979 and 1983 with the next two continuations.

From these dates, then, one finds that artificial intelligence has been portrayed in contemporary science

fiction without the dominance of any particular view. The rise of modern industry and its use of machines did not serve to strike fear into a majority of authors. Nor did the coming age of electronics quell the fears of many. In each decade some writers stressed the hazards of machine intelligence while others noted its potential benefits. As with most complicated issues, at no point does there seem to be a consensus among them.

BIBLIOGRAPHY

- Adams, Douglas. The Hitchhiker's Guide to the Galaxy. New York: Pocket Books, 1979.
- Anderson, Poul. Time and Stars. New York: Manor Books Inc., 1964.
- Asimov, Isaac. Asimov on Science Fiction. Garden City, New York: Doubleday and Company, Inc., 1981.
- . I, Robot. New York: Ballantine Books, 1950.
- . Lucky Starr and the Moons of Jupiter. Greenwich, Connecticut: Fawcett Books, 1957.
- . The Robots of Dawn. Garden City, New York: Doubleday and Company, Inc., 1983.
- Barnett, Jeffrey A. "An AIers Lament." Artificial Intelligence 6, 2 (1985): 54-56.
- Benford, Gregory. In the Ocean of Night. New York: the Dial Press, 1972.
- Clarke, Arthur C. "The Sentinel." In Expedition to Earth. New York: Ballantine Books, 1953.
- . 2001: a Space Odyssey. New York: Signet, 1968.
- . The Lost Worlds of 2001. New York: Signet, 1972.
- . 2010: Odyssey Two. New York: Ballantine Books, 1982.

- Conklin, Groff, ed. Science Fiction Thinking Machines: Robots, Androids, Computers. New York: Vanguard Press, 1954.
- "The Dark Ages of AI." Artificial Intelligence 6, 2 (1985): 122-34.
- Del Rey, Lester. "Helen O'Loy." In Souls in Metal. Ed. Mike Ashley. New York: St. Martin's Press, 1977.
- Dick, Philip K. Do Androids Dream of Electric Sheep. New York: Ballantine Books, 1968.
- Dunn, Thomas P. and Richard D. Erlich, eds. The Mechanical God Machines in Science Fiction. Westport, Connecticut: Greenwood Press, 1982.
- . Clockwork Worlds Mechanized Environments in SF. WestPort, Connecticut: Greenwood Press, 1983.
- Elwood, Roger, ed. Invasion of the Robots. New York: Paperback Library, 1965.
- Greenberg, Martin, ed. The Robot and the Man. New York: Gnome Press, 1953.
- Heinlein, Robert A. The Moon is a Harsh Mistress. New York: Berkley Publishing Corporation, 1966.
- Herbert, Frank. Destination: Void. New York: Berkley Publishing Corporation, 1966.
- . The Jesus Incident. New York: Berkley Publishing Corporation, 1979.
- . The Lazarus Effect. New York: Berkley Publishing Corporation, 1983.

- Hogan, James P. Code of the Lifemaker. New York: Ballantine Books, 1983.
- . The Two Faces of Tomorrow. New York: Ballantine books, 1979.
- Jones, D. F. Colossus. New York: Berkley Publishing Corporation, 1966.
- . Colossus and the Crab. New York: Berkley Publishing Corporation, 1977.
- . The Fall of Colossus. New York: Berkley Publishing Corporation, 1974.
- Knight, Damon, ed. The Metal Smile. New York: Belmont Books, 1968.
- Leinster, Murray. "A Logic Named Joe." In Souls in Metal. Ed. Mike Ashley. New York: St. Martin's Press, 1977.
- Moskowitz, Sam, ed. The Coming of the Robots. New York: Collier Books, 1963.
- Myers, Robert E., ed. The Intersection of Science Fiction and Philosophy: Critical Studies. Westport, Connecticut: Greenwood Press, 1983.
- Nolan, William F., ed. The Pseudo-People: Androids in Science Fiction. Los Angeles: Sherbourne Press 1965.
- Pohl, Frederik. Beyond the Blue Event Horizon. New York: Ballantine Books, 1980.
- . Gateway. New York: Ballantine Books, 1966.
- . Heechee Rendezvous. New York: Ballantine Books, 1984.

- Porush, David. The Soft Machine Cybernetic Fiction. New York: Methuen, 1985.
- Raike, William M. "The Fifth Generation in Japan." BYTE 10, 4 (1985): 401-06.
- Roddenberry, Gene. Star Trek: the Motion Picture. New York: Pocket Books, 1979.
- Rose, Frank. "The Quest for Thinking Machines." Science Digest 94, 5 (1986): 36-43.
- Ryan, Thomas J. The Adolescence of P-1. New York: Ace Books, 1977.
- Schank, Roger and Larry Hunter. "The Quest to Understand Thinking." BYTE 10, 4 (1985): 143-55.
- Simak, Clifford D. A Choice of Gods. New York: Ballantine Books, 1972.
- . City. New York: Ace Books, 1952.
- . "I Am Crying All Inside." In Souls in Metal. Ed. Mike Ashley. New York: St. Martin's Press, 1977.
- . Project Pope. New York: Ballantine Books, 1981.
- . Special Deliverance. New York: Ballantine Books, 1982.

APPENDIX

Works which place "man below machine":

The Adolescence of P-1, Thomas Ryan (1977)
 The Colossus series, D. F. Jones
 Colossus (1966)
 The Fall of Colossus (1974)
 Colossus and the Crab (1977)
 The Ship series, Frank Herbert
 Destination: Void (1966)
 The Jesus Incident (1979)
 The Lazarus Effect (1983)
Star Trek The Motion Picture, Gene Roddenberry (1979)
Time and Stars, Poul Anderson (1964)

Works creating "machine equality":

A Choice of Gods, Clifford Simak (1972)
City, Clifford Simak (1952)
Code of the Lifemaker, James Hogan (1983)
I, Robot, Isaac Asimov (1950)
Lucky Starr and the Moons of Jupiter,
 Isaac Asimov (1957)
Project Pope, Clifford Simak (1981)
The Robots of Dawn, Isaac Asimov (1985)
Special Deliverance, Clifford Simak (1982)

Works depicting "mechanical servitude":

2001 A Space Odyssey, Arthur C. Clarke (1968)
2010: Odyssey Two, Arthur C. Clarke (1982)
Do Androids Dream of Electric Sheep,
 Philip K. Dick (1968)
 The Gateway series, Frederik Pohl
 Gateway (1976)
 Beyond the Blue Event Horizon (1980)
 Heechee Rendezvous (1984)
Giants' Star, James P. Hogan (1981)
The Hitchhiker's Guide to the Galaxy,
 Douglas Adams (1979)
The Moon is a Harsh Mistress, Robert Heinlein (1966)