ARTIFICIAL INTELLIGENCE IN SCIENCE FICTION

BODY, MIND, SOUL—THE 'CYBORG EFFECT': ARTIFICIAL INTELLIGENCE IN SCIENCE FICTION

BY
DUNCAN LUCAS, B.A.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Arts

McMaster University

MASTER OF ARTS (2002) McMaster University
(English) Hamilton, Ontario

TITLE: Body, Mind, Soul—The 'Cyborg Effect': Artificial Intelligence in Science Fiction

AUTHOR: Duncan Lucas, B.A. (McMaster University)

SUPERVISOR: Professor Joseph Adamson

NUMBER OF PAGES: v, 185

Abstract:

Though this project is about representations of artificial intelligence (AI) in science fiction (SF), no discussion of 'artificial' intelligence could ever take place without considering 'real' intelligence. Consequently, at core and by default, this project is about human intelligence. Artificial intelligence throws into relief the essence of being human as a tripartite construction of body, mind, and their synergistic combination, by creating an intelligent, dialogical, and interrogative entity as a comparative Other.

Chapters one and two address two basic questions: What is science fiction? What is artificial intelligence? These evolve additional questions: How do science fiction writers delineate the physical and intellectual capabilities and capacities of humans versus machines (in the broadest sense), their comparative behaviours, and thereby, consider human methods for understanding our universe and our place in it? What place do SF writers imagine machine intelligence taking in our world? What are the ethical, moral, and social implications for human verus machine intelligence?

Chapters three and four consider how authors construct AIs, what physical forms they might take, and the relative importance of the body versus the mind. The imaginative creations are compared to actual develops in the science of AI, thereby revealing some surprising prophecies. Discovered is that human beings are nervous about their own technological constructions, especially when those creations begin to match human intellect and mentation. Consequently, some of our worst bigoted behaviours are brought the fore. I designate two temporal periods as the 'animated automaton' (citing *Frankenstein*, *Metropolis*, *R.U.R*) and the 'heuristic hardware' (*2001: a Space Odyssey*, *Do Androids Dream of Electric Sheep?*, *The Hitch Hiker's Guide to the Galaxy*, and *I, Robot*).

Chapter five focusses on a single narrative (*Galatea 2.2*) as an excellent consideration of the current state of AI research and the development of ever more effective systems for processing information. It begins with observing a change in the scientific worldview and the change from energy to information as the fundamental reality. The AI challenges a human to recognize and acknowledge humanity's own despotic and parodic behaviours. By considering exactly how human beings learn, know, and remember, it throws into dramatic relief our own assumptions about the superiority of human intelligence.

Chapter six looks at post-1980 literatures (*Neuromancer*, *The Matrix*, *The Diamond Age*, and *Terminator*) and the influence of the personal computer on the imaginations of SF writers. The narratives' complexities increase and the boundaries between assumed 'reality' and 'virtual reality' erode. Human beings are clearly anxious about increasingly powerful thinking machines, probably because our confidence in the uniqueness and singularity of human intelligence is challenged directly. The connection between body and mind is, paradoxically, both broken and affirmed, thus forcing humans to find ways to understand the essence of consciousness, particularly as it may related to the 'soul.'

Ultimately, AI could teach human beings about ourselves, and may force us to more clearly define 'human being.' Potentially, though not an expressed goal, AI research could unify humanity globally and help us to help ourselves in re-structuring economic, educational, social disparity.

Acknowledgements:

Dr. Alvin Lee, Dr. Joe Adamson, Dr. Jean Wilson for their consistent encouragement and mentorship — to say nothing of patience.
Dr. Imre Szeman, who assigned me the lowest undergraduate mark I ever received — and so raised the bar.
My thesis committee: Dr. Joe Adamson, Dr. Imre Szeman, and Dr. Anne Savage. — I hope it was worth it.
Dr. Don Goellnicht and Dr. Roger Hyman for helping me to believe in myself and my work at the graduate level, and encouraging my continuing studies.
To all my uncited undergraduate professors. If any of the ideas and information in this thesis are yours but which I fail to acknowledge, it is only because I learned so much that I no longer know where it all came from. Please take it as a compliment that I was listening on some level.
My family and friends, of course, for their unwavering support — and patience.
My friend, lover, partner, my psycho-emotional foundation, Trish O'Meara, for making home home by affirming and demonstrating the true value of human life — and her amazing patience. I love you, in body, mind, and soul.

Contents:

For man to tell how human life began
Is hard, for who himself beginning knew?
(John Milton)

Introduction: Where are we going?

We can only see a short distance ahead, but we can see plenty there that needs to be done.
(Alan Turing)

This project is about representations of artificial intelligence in science fiction. It begins with the assumption that an artificial intelligence has demonstrated, or inevitably will demonstrate, consciousness. My personal feelings on the likelihood of this scenario do not shape that statement; I accept the assumption because for the writers whom I consider in this thesis it is 'true' as a narrative premise.

No discussion of 'artificial' intelligence, however, can take place without considering 'real' intelligence. In an essay mixing "both computers and psychology," writes Marvin Minsky, the "reason is that though we'd like to talk about making intelligent machines, people are the only such intelligence we can imitate or study now" (in *The Age of Intelligent Machines* 219). At core and by default, then, this project is about human intelligence. From a human perspective, I will go a step further and suggest that the only conscious intelligence we may ever acknowledge and recognize as 'legitimate' is the one we want to recognize as uniquely intelligent — our own. In the perhaps not too distant future, the term 'artificial intelligence' may become an oxymoron, and we may be forced to drop the prefix 'artificial' when it becomes prejudicial and denigrating. But most people believe the future has not yet been written . . .

Writing an English Master's thesis with some understanding of literary principles is one practice, but the scientific principles behind artificial intelligence (AI) are another

discipline all together. Putting the two into a cohesive package requires exploration and synthesis based in the potential of free ranging thought associations. For this reason, I ask readers' indulgence and tolerance. I think this is an appropriate request in the science fiction context because the genre's readership, its 'fans,' are particularly accepting (forgiving? . . .) of ostensibly ludicrous extrapolations and speculations. Arthur C. Clarke has written, "Although it has become something of a cliché, perhaps the most important attribute of good science fiction—and the one that uniquely distinguishes it from mainstream fiction—is its ability to evoke the sense of wonder" (*Greetings* 404). Yet this project's very limitations force me to contain myself within a system, to provide boundaries and thus allow the expansiveness to work in tension with an equal and opposing force. Bridging the gap between art and science, or at least bringing art into dialogue with science, is a fundamental motivator behind this philosophical treatise, or speculation, or investigation, or creation, or analysis, . . . or whatever it be judged.

Does it not sometimes seem that Western society is steadily de-valuing literature's role in and importance to social development by placing too much emphasis on the Sciences, as if they alone could understand and define the essence of life and/or being human? Is our essence coded in our DNA and/or behaviour, or our thoughts, or elsewhere? Science may provide us with workable 'laws' for making technology, but in the end "we possess nothing but metaphors" ("On Truth and Lies" 83) as Friedrich Nietzsche so eloquently declared. The "clever beasts" (79) of Earth are a fortuitous, exploitive species alive on an obscure little marble dangling, spinning, and swinging on a Nothingness amid a vast Incomprehension. But are we accepting responsibility for our behaviours? Or are we

becoming increasingly arrogant about our knowledge, convincing ourselves that the mendacious is 'solid' and 'real,' and thus, we become increasingly ignorant? "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die out, and a new generation grows up that is familiar with it" (Max Planck as quoted in *The Great Thoughts* 333). Literature can not be subjected to scientific experimentation, but then neither can much of 'authorized' or 'official' theoretical physics, or human psychology. 'Good' scientists know this, implicitly understanding their task when they speak of 'describing' unseeable yet measurable physical properties. In his introduction to *Vehicles: Experiments in Synthetic Psychology* (1984), "Let the Problem of the Mind Dissolve in Your Mind," Valentino Braitenberg writes:

> This is an exercise in fictional science, or science fiction, if you like that better. Not for amusement: science fiction in the service of science. Or just science, if you agree that fiction is part of it, always was, and always will be as long as our brains are only minuscule fragments of the universe, much too small to hold all the facts of the world but not too idle to speculate about them. (1)

Literature's value is in helping to develop an understanding of what means 'human-being' (noun sense) *and* 'human being' (verbal sense). 'Human being' might be better thought of as not a thing, but an action, a becoming, a creating, a doing, a process. Artificial intelligence throws into relief the essence of being human by creating an intelligent and interrogative entity as a comparative Other without resorting to inventing gods as simply displaced and 'elevated' humans, nor anthropomorphically endowing the obviously unendowable inanimate.

For me, the humanities seem threatened by utter de-valuation within the current capitalist driven, political environment of the Western academy because it places too much

emphasis on business and technology. But the humanities may simply be in transition, in a re-valuing process. Humans continue to produce cultural, literary artifacts — films, novels, poems, and even video games with narrative underpinnings — literature itself being only one among many '*arti-types*.' Being contemporary with newly emerging works, we may not yet recognize their long term, individual import. Which of today's works will endure the 'test of time'? Alas, "*Ars longa, vita brevis*" (Hippocrates). In light of these artifacts, themselves so invested in creating a range of archetypes, we can conceive of and argue for the Imaginative itself as the value added.

Story telling must have fundamental human value because every person in every culture through all of history has told stories; we share anecdotes with friends; we recall lived events as narratives, a meaning-making process; we seek, in the remembering and telling of our daily stories, to structure and shape life cohesively. Whether we are 'right' or 'wrong' is surely less important that the process itself. Life is activity; value is added. Entertainment, then, is a social value, and there is no reason why entertainment can not be as instructive (perhaps even implicitly prescriptive) as it is meaning-making. Litterateurs are simply more sophisticated at imagining structures and meanings than most people, and more committed to committing their ideas to cultural remembrance. — Van Gogh was a better painter than most of us too.

At some level, we all engage in the scientific process everyday as we analyse, reflect on, and conclude from, sensually gathered, empirical data. This is the daily experience called living, a process made so mundane by its very constancy that it becomes a forgetting. Rational analysis and creativity are not separate (or separable) intellectual processes, of course, though they can be made discrete. They do and must work together. Curiously,

artificial intelligence researchers found mimicking human logic-thinking an easy problem to solve, while creative and associative thinking has frustrated them. Yet, in the wake of Kierkegaard's either/or distinctions, we often seem to accept individual thinking processes as either rational or not, and, in a culture too heavily biassed in favour of the rational, we easily dismiss as a non-value-added social exercise, the creative, the imaginative, the speculative. They *make* nothing. But they might describe anything and everything.

One time, standing puny and in awe of the aurora borealis, I remember hearing a nearby person ask, "What causes it?" 'Who cares?' I thought. 'Ask a poet.' Yet, even words failed the moment. Both scientific actuality — a particular particulate solar wind colliding with the atmosphere while bending around Earth's magnetism — and poetic expression were meaningless under the shadow of these seemingly audible northern spirit wisps, a duet of stardust and night, a visual song of shining shadows, Sirens singing of light and night, of magnificence and majesty and marvel. Magnetism or magic: Is there really a difference?

Part One: Conception.

One: What is Science Fiction?

> A new species would bless me as its creator and source;
> many happy and excellent natures would owe their being to me.
> (Mary Shelley)
>
> This is—*Hubris clobbered by nemesis.*
> (Brian Aldiss)

Science fiction is a 'slippery' term which might seem easily understood; upon reflection, however, it becomes difficult to define precisely. "If you have to ask what science fiction is, you'll never know" (anonymous quote used by Clarke, *Greetings* 398). According to my *Oxford Reference Dictionary*, science fiction is "a class of prose narrative which assumes an imaginary technological or scientific advance, portrays space travel or life on other planets, or depends upon a spectacular change in the human environment." This is not a good definition. Science fiction need not be prose, or even narrative, though these generic approaches have so far proved better suited to accomplishing writers' goals. Science fiction is dependant not on form but on content. Rhetorically, the definition implies a non-reality, an "imaginary" world by using the word "assumes" as if science fiction writers pluck possibilities *ex nihilo* from Imagination without necessarily considering their viability, plausibility, likelihood, or functionality. This is not so. Science fiction writers are especially interested in and informed about 'cutting edge' science and technology, and their major ideas are extrapolations from and speculations about the effects of scientific

discoveries and developments. The problem is, in part, indelibly connected to a vague distinction between fantasy fiction and science fiction. If you enter a bookstore, you will often find the singular marketing category 'science fiction *and* fantasy' as if these two genres can and should be lumped together.

Fantasy requires the acceptance of, belief in, working assumption and use of, what we call 'magic,' a character's agency and/or power to influence a physical world with purpose but without explanation, which is regarded as irrelevant or subservient to the author's desired effect-objective, the result. Fantasy does not care how 'it' is done, only that it is done and why it needs to be done. In 'reality,' we do not accept the notion that a person can simply snap a finger or wiggle a nose, or utter 'shazam,' to produce an actual, physical effect. The rules of physics do not (have to) apply in fantasy. J.R.R. Tolkien's *The Lord of the Rings* is a current popular, resurgent example.

Science fiction, however, assumes that character (usually human) agency has an explanation and executes power with deference to scientific principles. Even when positing a 'magical' *über*-character who implicitly or explicitly wields mysterious powers, or causes physical change and/or influence beyond 'realistic' comprehension levels, science fiction nonetheless implies that an explanation for the agency is not only possible but desirable by suggesting that the 'miraculous' powers circumventing the 'ordinary' laws of nature have a definite source in an accessible knowledge system, though the implied specific knowledge often remains assumed, postulated, or beyond the current database of a given epoch. Science fiction does not need to provide an explanation, only infer one exists. A character's power may appear magical when considered from a position of relative ignorance. If a twenty-first century person time travelled back to Plato's Greece with a lighter, might that

person not possess a 'magic talisman' which, for us, is a mundane technological device? Today, we have little trouble imagining a technological device, say a 'ray-gun,' though we have never held one. The essential technology exists; our imaginations can extrapolate the 'reality.'

Having delineated and separated science fiction from fantasy, I now restore the ambiguity with Arthur C. Clarke's 'third law of technology': "Any sufficiently advanced technology is indistinguishable from magic" (*Greetings* 413).[a] Further, given generic flexibility and the possibility of works embodying several categories (realism, romanticism, myth, gothic, etc.) simultaneously, a text may, at any time, shift its relative position, and a particular story can conceivably slide into the realm of science fiction as its existence marches into the future. We can conceive, then, of a range of possibilities for character agency with 'pure' fantasy, Northrop Frye's "romance mode" (*Anatomy* 33), defining one end of a scale and 'hard' science (high/low mimetic mode) defining the other end.[1]

In 1816, Mary Shelley published the first genuine SF novel, *Frankenstein*. Her novel is (has been labelled) fantasy, gothic, 'gothic romance,' horror, social satire, but it is not often referred to by the 'science fiction' appellation, except by SF admirers. Mind you, they are probably most qualified to decide. In *Trillion Year Spree: the History of Science Fiction*, Brian Aldiss[2] observes that at the time of publication, the "division between the

---

[1]Taking my lead from Brian Aldiss, I shall henceforth use the abbreviation 'SF.' "That down-market appellation 'sci-fi', sometimes heard on the lips of the would-be trendy in the media and elsewhere, is purposely avoided. We bow to the fact that much of what passes for science fiction these days is nearer fantasy. SF can, after all, be imagined to stand for science fantasy, as it can for speculative fiction (for those who are attached to that term)" (*Trillion Year Spree* 20).

[2]A competent SF writer himself, Aldiss' excellent and comprehensive (and substantial) volume maps major contributions to SF, including a detailed accounting of Mary Wollstonecraft Shelley's (nee Godwin) influences, and the literary, scientific, and intellectual atmosphere of her time, place, and experience.

arts and sciences had not then grown wide" (30). Talking about the "uniquely innovative" (39) features of *Frankenstein*, he writes:

> Interest has always centred on the creation of the nameless monster. This is the core of the novel, an experiment that goes wrong—a prescription to be repeated later, more sensationally, in *Amazing Stories* and elsewhere. Frankenstein's is the Faustian dream of unlimited power, but Frankenstein makes no pacts with the devil. 'The devil' belongs to a relegated system of belief. Frankenstein's ambitions bear fruit only when he throws away his old reference books from a pre-scientific age and gets down to some research in the laboratory. This is now accepted practice, of course. But what is now accepted practice was, in 1818, a startling perception, a small revolution. (39-40)

Aldiss also points to a fundamental requirement in the scientific process, the authentication and confirmation of a discovery through an experiment's repeatability: "As if to dispel any doubts about her aversion to 'jiggery-pokery magic', Mary makes it plain that her central marvel shares the essential quality of scientific experiment, rather than the hit-and-miss of legerdemain. She has Frankenstein create life a second time" (41).

Mary Shelley's ideas are not *ex nihilo* creations, but belong to the intellectual speculation in her time: "Among his other capacities, Erasmus Darwin [grandfather of Charles Darwin] was a copious—and famous—versifier. In his long poems he laid out his findings on evolution and influenced the great poets of his day" (Aldiss 30); and, from Mary Shelley's own preface to her novel: "The event on which this fiction is founded has been supposed, by Dr. Darwin and some of the physiological writers of Germany, as not of impossible occurrence" (*Frankenstein* xiii). In her introduction, she explicitly indicates her *intent* was to write a 'ghost' story, yet scientific knowledge and experimentation are a vital impetus in the tale. Victor Frankenstein's skills are presented as those of a scientist, or more accurately a 'natural philosopher,' pursuing discovery. His actions are posited as empirical.

The litmus test for science fiction is whether or not one can remove reference to science and/or explanation of agency without altering the essential story. With *Frankenstein*, this is not possible. *Frankenstein*, therefore, is science fiction.

Frankenstein marks not only the beginning of SF but, relevant to the purposes of this project, it is also antecedent to the science broadly termed 'artificial intelligence.' Shelley was, after all, representing the 'manufacturing' of a sentient being, a thinking, nameless entity capable of saying, "I." Up to this point, narratives dealing with the artificial creation of 'humanoid' life, or imitations of humans and their behaviours, paradoxically belong in the religious-literary tradition and mystical realm as with the Golem, the animated clay figure in Hasidic Jewish mythology, and Ovid's Pygmalion and Galatea tale in *Metamorphosis*.[3] Shelley's fundamental differentiation is animation not by magic or the supernatural, not by the gods, but by a man — selfishly, hopefully, foolishly, blindly: "The 'vital spark' is imparted to the composite body. Life is created without supernatural aid. Science has taken charge. A new understanding has emerged" (Aldiss 40). That new understanding, so familiar to us today, is the ability to manipulate the 'natural order' purposefully in the interest of satisfying humans' curiosities and multifarious desires.

As this project evolves, the full implications of Shelley's 'vision' will become clearer and, I believe, demonstrate SF's primary and vital social function in our technology driven culture: the speculative and experimental testing of scientific principles extrapolated to a time of crisis.

---

[3]See Samuel Holmes Vasbinder's excellent account, *Scientific Attitudes in Mary Shelley's Frankenstein*, particularly chapter four, "Early Literature on Artificial Humans."

Two: What is Artificial Intelligence?

AI is the study of how to make computers do things at which, at the moment, people are better. (Elaine Rich)

Artificial Stupidity (AS) may be defined as the attempt by computer scientists to create computer programs capable of causing problems of a type normally associated with human thought. (Wallace Marshall)

With one small mechanical invention, the flying shuttle, Western civilization took a giant technological leap in the 18th century; the industrial revolution began. In *The Age of Intelligent Machines* (1990),[1] artificial intelligence researcher and entrepreneur Raymond Kurzweil suggests that the computer age is a second industrial revolution. The distinction is simple: "The Industrial Revolution of the last two centuries—the *first* Industrial Revolution —was characterized by machines that extended, multiplied, and leveraged our *physical* capabilities. . . . The *second* industrial revolution, the one that is now in progress, is based on machines that extend, multiply, and leverage our *mental* abilities" (*Intelligent* 7).

Computing, as we understand it today, was initiated by Charles Babbage in 1821 with his invention, the Difference Engine, and its later refinement, or 'evolution,' the Analytic Engine (AE). He never completed its construction, however. Historically

---

[1]This essay owes a large debt to Kurzweil's book. Over five hundred pages long, it summarizes and overviews the AI industry as of 1990, the first seventy pages alone dealing with the philosophical foundations, particularly around issues of language, communication methods, and word definitions, plus human psychology, thought and emotion, plus again theoretical physics, mathematics and electronics. Reading it proved to be a 'mind bender.'

significant as the first 'programmable' device using "a punched-card reader inspired by the Jacquard looms, automatic weaving machines controlled by punched metal cards" (Kurzweil, *Intelligent* 165), the AE would have been capable of carrying out logical computation. Programming itself was developed by the daughter of poet Lord Byron, Lady Ada Lovelace, who is responsible for inventing "the programming loop and the subroutine" (167).[2]

The term 'artificial intelligence' (AI) was coined in 1956 at the Dartmouth Summer Research Project on Artificial Intelligence. Previously, in 1950, British mathematician Alan Turing explored and explained many of the mathematic conditions, with large philosophical implications, for nascent AI research in his article "Computing Machinery and Intelligence." Most importantly, Turing proposed what is now known as the 'Turing test.' *The* theoretical test for machine intelligence, it is an elegant and simple exam. (And I would wager not a few humans are capable of 'failing' this unfailable test.) He asks a simple question: "Can machines think?" Immediately recognizing the question's problematic dependence on definitions for 'machine' and 'think,' he writes: "Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words" (Turing 433).

He called the new problem the 'imitation game': One man (A) and one woman (B) are placed in a closed room. Another man (C), the 'interrogator,' communicates with A and

[2]The "only legitimate child of Lord Byron, the poet . . . Ada Lovelace is regarded as the world's first computer programmer and has been honored by the United States Defense Department, which named its primary programming language, Ada, after her" (Kurzweil, *Intelligent* 167).

B using a teletype machine only (to mask their voices) and may ask *any* question(s).[3] C must now determine which of A or B is the woman. A, however, is trying to fool C into choosing him as her. B's objective is to help C confirm her identity. How long will C take to correctly identify B as the woman? Having posited this scenario: "We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'" (Turing 434). As of 2002, no machine has come close to passing this test, except in literature. In Arthur C. Clarke's *2001: A Space Odyssey*, HAL "could pass the Turing test with ease" (*2001* 97).

Since its 1950 publication, the Turing test's basic premise has been abused at times when scientists attempted to *limit* the range of questions. Turing's intent allows any and all questions. But Turing placed his own limitations on the game; he allowed that only digital computers would play. In simplified form, only one person and one computer need be involved. If a computer can sustain a dialogue sufficiently to 'fool' the interrogator, the machine 'wins.' According to philosopher Daniel Dennett, "Turing proposed that any computer that can regularly or often fool a discerning judge in this game would be intelligent, a computer that thinks, *beyond any reasonable doubt*" (in *The Age of Intelligent Machines* 48). Novelist Richard Powers writes: "A perfect, universal simulation of intelligence would, for all purposes, *be* intelligent" (*Galatea 2.2* 52). However, Turing asks: "May not machines carry out something which ought to be described as thinking but which

[3]With the advent of speech recognition and synthesis, a fully aural/oral and dynamic approach to the game became possible and practical.

is very different from what a man does?" (435).

As he challenges various points of view and opinions on the viability of thinking machines, Turing begins by stating his own position and a prediction for the future:

> I believe that in about fifty years' time it will be possible to programme computers . . . to make them play the imitation games so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning. The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (442)

He was right. In *The Age of Intelligent Machines*, Kurzweil, based on a survey of six children ages seven to nine as "naive experts," concludes that "computers, or at least the computers that these children have had experience with, are not conscious, but they do *think*, and therefore thinking does not require consciousness" (39).[4] We are left, then, with ambiguous definitions for words like 'think' and 'conscious,' the two words most likely to support 'intelligence.'

When I began this project, I vaguely believed that a basic synopsis could be articulated for intelligence. I read multiple viewpoints on the subject, and lost confidence in that belief. Philosophers, natural philosophers, scientists, even engineers, have been debating (or arguing, usually politely) the nature of 'intelligence' for a long time, a really long time. (Ah, but a really long human time is a universal 'hiccup' . . .) I turn once more

---

[4]Six questions inform this conclusion: "Can a computer remember?" Yes. "Does a computer learn?" Yes. "Do computers think?" Two negatives, four affirmatives, because they sometimes take time to respond to commands. "Do computers have feelings?" Unanimously, 'No,' followed by laughter at the absurdity of the suggestion. "Do you like computers?" Yes. "Do computers like you?" This question was dismissed as irrelevant (Kurzweil 39). As Kurzweil rightly points out, these questions say as much about children's understanding of the words 'think,' 'feel,' 'remember,' as they do about their attitudes toward machine intelligence.

to Alan Turing:

> The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration. If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence. With the same object, therefore, it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behaviour. (Alan Turing in 1947, as quoted in *Age of Intelligent Machines* 14)

I am not interested in a semantic debate. But, if thinking is a collection of intelligent processes, and if thinking does not require consciousness, then intelligence may not require consciousness either, though consciousness may require intelligence. That remains to be determined.

"The reader must accept it as a fact that digital computers can be constructed, and indeed have been constructed, according to the principles we have described, and that they can in fact mimic the actions of a human computer very closely" (Turing 438). This is why Turing, in recognizing human thinking as comprised of a range of discrete processes, focussed on human style communication. Fundamentally, the Turing test is not simply interrogative, but dialogical. Mikhail Bakhtin describes dialogue as the only viable method for determining human 'truths,' or ideas: "The idea . . . is not a subjective individual-psychological formation with 'permanent resident rights' in a person's head; . . . The idea is a *live event*, played out at the point of dialogic meeting between two or several consciousnesses" (*Problems* 88).

In *Frankenstein*, some of the most compelling scenes (for me) are the dialogues between the Being and his maker, particularly his self-narrated life-story. If one accepts the Being as a type of AI (further proof to come later), then he easily passes the Turing test.

When creator and created meet for the first time, Frankenstein exclaims:

> 'Devil . . . do you dare approach me? And do not you fear the fierce vengeance of my arm wreaked on your miserable head? . . .'
> . . . 'I expected this reception,' said the demon. 'All men hate the wretched; how, then, must I be hated, whom am miserable beyond all living things! . . . Be calm! I entreat you to hear me before you give vent to your hatred on my devoted head.' (Shelley 95).

I have edited this scene to highlight both the dialogical and interrogative quality of this 'live event.' The Being clearly 'thinks' and demonstrates 'intelligence'; he is clearly conscious, or self-reflective. But he has also behaved according to the worst human capability, willful and 'demonic' destructiveness. Frankenstein's Being reveals that 'intelligence' carries no moral imperative.

The 'imitation game' creates a "fairly sharp line between the physical and the intellectual capacities of a man" (Turing 434).[5] However, it "might be urged that when playing the 'imitation game' the best strategy for the machine may possibly be something other than imitation of the behaviour of a man" (435). Part of, my purpose, then, will be to consider how SF writers delineate the physical and intellectual capabilities and capacities of humans versus machines (in the broadest sense), their comparative behaviours, and thereby, consider human methods for understanding our universe and our place in it. What place will machine intelligence take in our world? What are the ethical, moral, and social implications of human verus machine intelligence?

What, then, is intelligence? Why do humans (seemingly) assume exclusive rights to

---

[5]In the context of literary studies some critics may wish to interrogate Turing's gender specificity. In purely scientific terms, that is irrelevant. As for his statement, "intellectual capacities of a *man*," we must move beyond (though not dismiss) this worry, regarding it as a 'reflection of his time' or personality.

the processes we call thinking? Why has thinking conventionally carried automatic inference of consciousness? Or does it? How and why does intelligence infer consciousness? Or does it? Can intelligence be measured through behavioural analysis? Whose behaviour sets the paradigm?

When we disassemble human thinking into constituent parts, or at least discrete types of processes, as AI researchers have done, we discover that previously mystifying thinking processes are no longer mysterious and that they are machine replicable — some of them. According to Marvin Minsky, "'intelligence' seems so elusive [because] it describes not some definite thing but only the momentary horizon of our ignorance about how minds might work" (*Intelligent* 214). Kurzweil makes an important distinction: "If we can replace the word 'artificial' with 'machine,' the problem of defining artificial intelligence becomes a matter of defining intelligence. . . . a process of learning, reasoning, and the ability to manipulate symbols" (*Intelligent* 16). Kurzweil makes a compelling argument (though Minsky disagrees) for evolution as "the ultimate in intelligence—it has created designs of indescribable beauty, complexity, and elegance. Yet, it is considered to lack consciousness and free will—it is just an 'automatic' process" (20). An intelligent process, however, is quite different from the abstract we call 'intelligence' which, as Kurzweil implies, does seem to imply and require consciousness and/or free will. Human intelligence requires intent as the control for a thinking process, or the considering and choosing of options to satisfy a designated or identified need.

"Evolution has achieved intelligent work on an extraordinarily high level yet has taken an extraordinarily long period of time to do so" (Kurzweil, *Intelligent* 21). In *The Age of Spiritual Machines* (1999), Kurzweil calls evolution an intelligent process because, in a

chaotic universe, 'natural selection' (even before the appearance of organic life) is sufficient to overcome the "second law of thermodynamics, sometimes called the Law of Increasing Entropy" (12), but only just. At a universally ponderous pace, and for some unknown 'reason,' molecular complexity gradually increased, but on that universal scale "the order represented by the existence of life-forms is insignificant in terms of measuring overall entropy" (*Spiritual* 13). He points out that evolution "is a process, but it is not a close system. It is subject to outside influence, and indeed draws upon the chaos in which it is embedded" (*Spiritual* 13). In short, on a linear time scale, life emerged from chaos. He makes this argument in order to suggest that technology follows, like organic life, an evolutionary process; this is the business of 'R & D' — research and development. Human intelligence is 'better' than evolution because it can make selections in a timely manner.

If we conceive technology as applied knowledge, then knowledge itself must also develop cumulatively. Assuming the 'big bang' gave rise to the universe, that moment instituted the physical 'laws,' the scientific study of which we call physics. Physics leads to chemistry; chemistry in turn leads to biology. "*A key requirement for an evolutionary process is a 'written' record of achievement*, for otherwise the process would be doomed to repeat finding solutions to problems already solved" (Kurzweil, *Spiritual* 13) [orig. emph.]. In bio-evolution, DNA is that record. Like nature, human intelligence is successful because it remembers; and better than nature because it keeps the record intentionally, knowingly.

The assignment of the appellation 'intelligent' to another entity is largely intuitive, and similarly, recognizing another intelligent entity as 'conscious.' We just 'know.' We willingly discuss animals (dogs and dolphins, for example) in terms of relative intelligence, but we balk at calling them conscious, at least in the way we understand conscious as a

reflection of self-awareness. We do not absolutely know, however, that animals are not self-aware; we do not dialogically interact with creatures other than humans. In compromise, we might call them sentient. Sentience seems to be a composite of (at least partial) consciousness and (at least partial) intelligence; sentience is feeling and perceiving. We look directly into the eyes of an Other and 'know' they are 'in there,' perceiving us in return, including animals as we exchange gaze with our pets and believe that they are sentient (empathetic, emotional . . . loving); we feel their affective presence. Human intelligence involves empathy.

Artificial Intelligence is *not* a type of life, artificial or real,[6] but the replication of (not necessarily exclusive) human thinking processes by an object of human construction. In 1990, Kurzweil suggested that the "human race, then, may very well be smarter than its creator, evolution" (*Intelligent* 21). Nine years later, he asks: "Can an intelligence create another intelligence more intelligent than itself? (*Spiritual* 40). In one way or another, all the narratives I will discuss are underwritten with that question. If the answer is 'no,' we have no cause for anxiety. If the answer is 'yes,' . . .

---

[6]At the risk of ambiguity and confusion, there is a type of computer programme call 'artificial life.' "Artificial Life: A sequence of outputs produced from a computer *program* that are presented with an initial configuration of points (the 'organism') and a set of rules (the 'genetic code') to generate subsequent generations of the organism. Artificial life is modeled on evolution by natural selection. Certain initial configurations and rules can produce visually pleasing images. This is thus one technique for generating computer art" (Kurzweil, *Intelligent* 541).

a. a.

Arthur C. Clarke's "Three Laws of Technology":

1. When a scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.
2. The only way of discovering the limits of the possible is to venture a little way past them into the impossible.
3. Any sufficiently advanced technology is indistinguishable from magic.

Part Two: The Birth of Artificial Intelligence in Narrative.

Three: The Animated Automaton.

> From Platonic times, inventors were anxious to apply whatever technology was available to the challenge of re-creating human mental and physical processes.
> (Raymond Kurzweil)

> If one read them the *Encyclopedia Britannica* they could repeat everything back in order, but they never think up anything original. They'd make fine university professors.
> (Karel Čapek)

The earliest technological devices, sticks and stones and bones, supplemented the physical capabilities of our evolutionary ancestors. These were, essentially, mechanical aids, tools to help procure food and to defend against the predators early hominids were poorly equipped to resist. And yet, not only did they battle many beasts, they ultimately won through the power of tools and the thoughtful application of tools to specific tasks. Over time, granted a long period of human time, the range of devices for mechanical assistance expanded, but always in the primary service of our physical abilities. As far back as Plato's Greece, inventors were building mechanical devices capable of imitating human behaviours, as, for example, an entirely mechanical orchestra built in China around 350 BCE. By the eighteenth century, steadily improving skills in miniaturization and watch-making led to P. Jacquet Droz's 1772 construction the "automaton L'Écrivain" (Kurzweil, *Intelligent* 160) which was capable of mimicking an explicitly human behaviour, writing continuously with a real pen. Two behaviours seemingly intellectual in essence, making music and writing, became mechanized, body-centric exercises.

We might be tempted to dismiss *Frankenstein* as a representation of AI because it diverges from what we initially tend to thing of as AI, the electronic computer. As I mentioned previously, AI is not a type of life, artificial or real, but the replication of discrete thinking processes by a device of human construction. So far, many of those processes have been successfully replicated using mechanical, electrical, or electro-mechanical devices. Frankenstein's creature is a living organism, a 'mammal,' arguably even an 'animal.' There is nothing inherent in AI or computer technology requiring electronics. The first computers were mechanical; they did not become electrical until science and technology acquired sufficient control over electro-magnetic and electromechanical energy to make them possible because before "the taming of the electron, this meant harnessing the state of the art in mechanical techniques" (Kurzweil, *Intelligent* 159):

> The fact Babbage's Analytic Engine was to be entirely mechanical will help us to rid ourselves of a superstition. Importance is often attached to the fact that modern digital computers are electrical, and that the nervous system also is electrical. Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance. Of course electricity usually comes in where fast signalling is concerned, so that it is not surprising that we find it in both these connections. In the nervous system chemical phenomena are at least as important as electrical. (Turing 439)

We have, then, a mechanical-electrical-biological continuum of body-centric, physical behaviours (at the atomic and cellular level), with corresponding scientific knowledge and technologies.

Now, in 2002 CE, the artificial construction of carbon based, organic entities is not simply a fantastical speculation but a conceivable practice. Bio-technology was a political and philosophical 'hot topic' in 2001 — the year the first human stem cells were cloned —

but of potentially more social significance for the west, because not an abstract but a direct business impetus to develop technologies for profit in spite of ethical concerns, bio-tech capitalization stocks were 'hot' investment commodities. 2001 is also the year when Celera Genomics Corp., taking credit for deciphering the human genetic code, published the first (almost complete) human genome. Mary Shelley, and later, Karel Čapek (pronounced show-peck), though unaware (who can foresee the future?) of what science would discover and uncover decades, even centuries later, 'foresaw' one of today's hottest research areas, genetic engineering.

With *Frankenstein,*[1] *Metropolis*, and *R.U.R.* (*Rossum's Universal Robots*), Shelley, Fritz Lang, and Čapek leap far into the (as yet unrealized) future of AI while maintaining and sustaining an immediate interest in and respect for both the futurist imaginings of their day and their immediate social conditions. With a history of accelerating scientific discovery and technology in the nineteenth and early twentieth centuries, Mary Shelley's mythopoeic legacy, and their social conditions feeding their imaginations, Čapek and Lang (with Thea von Harbou) envisioned essentially mechanical beings, but with a significant organic composition, or what I am calling the 'animated automaton.' Let me influence that statement's interpretation. By 'beings' I infer a verbal understanding, types of behaviours, types of doings, but always with an ambiguous sub-inference to a physical, corporeal existence, a body. Thus, we can envision human-like bodies performing multiple actions in mechanical ways, or automatically.

---

[1]While I pinpoint *Frankenstein* as an early intersection of AI and SF in literature, there are other examples of writers representing automatons, such as E.T.A. Hoffmann's short stories "Automata" (1814) and "The Sandman," published in 1816, the same year as *Frankenstein* (1816). See Vashbinder for "The Literature of Artificial Humans Prior to 1818," chapter four of *Scientific Attitudes in Mary Shelley's Frankenstein*.

In discussing AI in SF, artificially constructed life, as antecedent to robotics, must be represented, particularly in the context of *Frankenstein*, and later literary developments such as Lang's *Metropolis* and Čapek's *R.U.R.*, because the "field of robotics is where all of the AI technologies meet: vision, pattern recognition, knowledge engineering, decision-making, natural-language understanding, and others" (Kurzweil, *Intelligent* 320). The conceptual unification of the mechanical automaton with the electronic computer is what we today call the robot (though robots in no way need be humanoid in form). Lang and Čapek could not have known this, however, because the 'field of robotics' had not yet been created, though Čapek, in fact, coined the term 'robot,' though his sense of the word more closely aligns with what we now call the 'android.'[2]

These seminal texts are invested with a 'god-complex': if a god will not (con)descend to our earthly level, or if a god can not be pulled down, why not raise humanity to become the god? Mythologically, they are all associable with God's inspiration of Adam in Genesis, the Pygmalion and Galatea narrative in Ovid's *Metamorphosis*, and the Golum of Hasidic Judaism. In the burgeoning SF context, however, divine intervention is eliminated and agency lowered to a human level and perspective with an implied and/or presumed science. Čapek's character, Old Rossum, invents a "protoplasm" (Novack-Jones translation 38), a 'primordial ooze' from which organic beings are manufactured, though with a less complex organization than humans. In Lang's *Metropolis*, the robot begins as a

[2]'Robot' comes from the Czech word 'robota,' meaning "forced labour" (*OED*), and enters into English usage when Čapek uses it in his short story "Opilek" (1917), and, more importantly, in the stage play *R.U.R.* (1921) (Kurzweil, *Intelligent* 312). However, "Karel Čapek was apparently not the inventor of the term 'robot;' he gives credit for that to his brother Josef" (Kussi 33). Kussi translates 'robota' as 'heavy labour.' Isaac Asimov defines it as "compulsory labor" (*Caves of Steel* vii).

mechanical device which is subsequently transformed into a biological entity through an alchemical-like — 'magical' — matter transformation. Frankenstein assembles his creation from existing and found body parts which he grafts together. 'Life,' or animation, is imparted by electric shock.[a] At core, each is a consideration of human social valuation and responsibility, that is, the value of human life for human beings and their responsibilities to that life.

Frankenstein did not necessarily 'do wrong' in making his creation; his was a post-creation failure of accountability. As critic Robert Wexelblatt suggests, "Shelley's novel is our first and still one of our best cautionary tales about scientific research; it is the literary and philosophic equivalent to the crude Luddite reaction to industrialization. The issues of *Frankenstein* are no different, basically, from those around which public debate on nuclear power, pollution, and genetic research are now centered" (Wexelblatt 116). Blinded by his well criticized Promethean ambition, Frankenstein's attitude in manufacturing a new life form has an uncanny ring compared with objections to bio-tech today. Analogously, he articulates people's specific fears regarding genetics and cloning in his arrogant assumption of power and ability to control that power: "A new species would bless me as its creator and source; many happy and excellent natures would owe their being to me" (Shelley 52). This is the 'god-complex.' Typified by Aldiss as 'Faustian,' with his "dream of unlimited power" (Aldiss 39), Frankenstein isolates himself in his attic laboratory while creating 'Being' and discovering the power to impart life. After animating his creation, he subsequently isolates Being by abandoning his parodic progeny. He runs away in horror

rather than face the ugliness for which he is explicitly accountable, Being's body not being entirely human in origin but assembled from parts scavenged in the "dissecting room and the slaughter-house" (Shelley 53). Often, claims Frankenstein, did his "human nature turn with loathing from [his] occupation, whilst, still urged on by an eagerness which perpetually increased" (53), he nonetheless persists. Frankenstein obsesses on the practical application and achievement of his theory at the cost of (self-)reflection on the possible ramifications. When finally confronted, face to face, by his now matured 'child,' he admits only a vague awareness of his culpability: "For the first time, also, I felt what the duties of a creator towards his creature were, and that I ought to render him happy before I complained of his wickedness. . . . I consented to listen, and seating myself by the fire which my odious companion had lighted, he thus began his tale" (97).

"Like Adam," says Being, "I was apparently united by no link to any other being in existence; but his state was far different from mine in every other respect. He had come forth from the hands of God a perfect creature, happy and prosperous, guarded by the special care of his Creator; he was allowed to converse with and acquire knowledge from beings of a superior nature, but I was wretched, helpless, and alone" (124). In most moral readings of the text, Being embodies evil. However, he can be characterizes as a 'victim' of social conditioning. Clearly, his behaviour is despicable in destroying innocent lives without a genuine motive; his actions are aggressive, not defensive. He selects victims for execution though they have personally done nothing to him and are simply associated with Frankenstein. He goes 'too far.' However, he is right to be angry. If we accept him as a sentient being, we can sympathize with that anger. What does he want? To live in society,

to have companionship, someone with whom to speak, to be loved, to have community and attachment. This is a basic human capability and necessity, what Martha Nussbaum calls "affiliation" ("Human Capabilities" 78). And yet, despite his vague intuition of responsibility, Frankenstein says, "There can be no community between you and me" (96). In murdering his maker's personal community and familial affiliations, Being is striking back at precisely the emotional place where Frankenstein hurt him — the 'heart.'

When Frankenstein's own sense of shame leads him, during Being's construction, to shun "my fellow creatures as if I had been guilty of a crime" (55), that same shame sense causes him to shun the newly animated Being whom he does not even consent to name. He imposes his own misanthropy on Being whose first social contact, then, amounts to a body based shaming and a parent's rejection. Being subsequently finds that he is utterly repulsive to the broader society as he meets repeatedly with prejudice. In terms of emotional development, all interpersonal contacts (counterpointed by one almost successful relationship with a blind man) lead him to intense and persistent feelings of shame. This constant shaming produces, inevitably, contempt, "a form of anger in which we declare the other person, this object of our negative affect, as far beneath us and worthy only of rejection. The purpose or function of contempt seems to be to instill in the other person a sense of self-dissmell or self-disgust and therefore shame at self-unworthiness" (Nathanson 129). Judged morally, Frankenstein's behaviour is initially the more reprehensible and shameful.

Alone in an alien world, Being's early psycho-emotional development is through a continuous social persecution which trains his emotions parodically, thus disposing him to

behave badly; essentially, he has a mis-guided early childhood education. If every social situation is a negative experience, then, as affect theory predicts, any subsequent similar environmental (in its broadest sense) condition or stimulus, any matching emotional pattern, will automatically trigger at least one of the bio-chemical processes fear-terror, distress-anguish, anger-rage. Consequently, Being evolves into a sociopath, a type of person who is "chronically unable to sustain any form of authentically intimate relationship" (Nathason 350). At their first genuine meeting, Being says, "I was benevolent and good; misery made me a fiend" (96).

In psycho-experiential terms, he is a victim of social persecution, prejudice. Not only is he subjected to ocular discrimination (a disgust trigger), he is judged by a moral code unknown to him. He is never educated or informed about the rules of social behaviour or the boundaries of proper actions. He is thrust into a wild orphan's life, without communal or familial guidance, without support.[b] Being, in murdering, simply shows what he has learned about the value of life from living, as opposed to a pedagogical and/or theoretical inculcation, from Frankenstein, and from the people with whom he later interacts. Lives are but the toys of self-serving gods, and he tells Frankenstein as much during their first *tête-à-tête*: "How dare you sport thus with life?" (95). Being is human, and yet he is not human. He talks like a human, but the subsequent violent actions are excessive as, driven by rage, he becomes a predator. We can not sympathize with his decision to act in this way. Unlike an inanimate machine, as for example the relentless, unfeeling 'cyborg' in James Cameron's film *Terminator*, Frankenstein's Being has an intensely complex emotional life. He is super-human, not only in physical stature, but in emotional stature as

well. He is capable of self-reflection and empathy, but he rejects the latter.

Being demonstrates a superior, almost too easy, ability to learn. But Shelley is less concerned with exactly how he develops than with the ramifications of Frankenstein's actions.[c] Frankenstein's Being defines our first coordinate on the AI spectrum, the artificially constructed, too human human. He represents the biological function as an emotional, operational modality with life imparted by an electrical spark. He represents, also, the conceptual shift from the origin of human beings as a divine act to one that is an unexplained mystery but nonetheless replicable by human action. Shelley, aware of Erasmus Darwin's early speculations on evolution (later developed and refined by his grandson, Charles Darwin), puts the power of life into human hands. (By contrast, humans have always had the power of death in hand.) Thus, the clay figures of mythology have, with *Frankenstein*, transmuted into artificial, organic constructs. In the end, I am left with an unanswered question: while he does listen to Being's story, thereby seeming to establish a rudimentary communal tie, by what lack of humanity is Frankenstein unable to empathize with Being and agree to manufacture him a mate with genuine concern and good will?

In Lang's *Metropolis* (1928), AI is represented in a feminine 'Franken-Being,' the Maria robot (Maria-R), a mechanical machine 'mysteriously' made animate and organic through an electrically powered process. Maria-R's life is due not to miracle and mystery but to the manipulation of physics, posited as pseudo-fact rather than fantasy. There may be no 'real' scientific application making this specific event 'true,' but, as SF requires not reality but implied human control, this poses no problem. As a viewing audience, we accept

the assumed science in order to engage the more important narrative and thematic issues.

When technocracy ascends the political spectrum to a position akin to religious fervour, the value of human life becomes subservient to the value of technology. I personally believe that the anxiety associated with the technology driven culture in *Metropolis* only increased through the twentieth century and is figured by the current neo-hippy, anti-globalization and nascent "neo-Luddite" (Kurzweil, *Spiritual* 197) movements. Humans need to feel they are making an individual contribution to the communal body or risk disenfranchisement which is a common theme in Modernist literature. Thus, by usurping their labour and thereby threatening the workers' necessity to the greater community, technology threatens to take away their human lives' purpose and sense of self-worth. In van Harbou's original novel *Metropolis*, Freder says, "And near the god-machines, the slaves of the god-machines: the men who were as though crushed between machine companionability and machine solitude. . . . They have no loads to carry: the machine carries the loads. They have not to lift and push: the machine lifts and pushes. They have nothing else to do but eternally one and the same thing, each in his place, each at his machine" (Kracauer translation 39).[3] What emptiness of life that humans might live in servitude to the machines— a suffering worthy of Sisyphus: "The man before the machine was no longer a human being. Merely a dripping piece of exhaustion, from the pores of which the last powers of volition were oozing out in large drops of sweat" (Kracauer 51). Though the machine does the work, it is the human who is exhausted, drained of energy. Freder's concern and compassion for an exhausted man leads him to relieve the man from

[3]For clarity and ease, I take a number of excerpts from an article by Siegfried Kracauer because it includes translated sections from Thea von Harbou's original novel.

working with the machine and, thereby, he experiences that immense fatigue directly: "Title: Father, father — I did not know that ten hours can be torture." (The *mise en scène* next shows the 'saviour' Freder in parody crucifixion, impaled with arms spread against the clock-like machine controls.) Recognizing 'will' as distinctly human, machines lack will except to the extent that people give them instructions so that they appear to do things only due to humanly imposed purpose. Yet, working with machines is 'sucking the life out' of the people, bleeding them of their volition, their will, in other words, turning them into machines, automatons.

Given that the American release edited out about seven of the original seventeen reels of *Metropolis*, Paul M. Jensen analyses the novel to highlight problems with the film and provide important connections and explanations concerning character motivations and their behaviours. Why does Jon Frederson, for example, want to incite worker violence? With the consequent destruction of their homes and the machines, "his method cripples the city's ability to function," so that "he is also working against his own interests and those of the upper classes he represents" (Jensen 7-8). He continues, "[t]hough both film and book are philosophically muddled, it is still possible to isolate certain themes. For example, the duality of human nature that fascinates Lang is here in abundance. The split in each individual between the mental and the physical has evolved, by the year 2000, into a social division. One group of people retains only the brain, while another uses only muscle" (Jensen 10). *Metropolis*, then, is fully invested in separating body and mind into discrete sites of action in order to critique, emphasize, and understand the inherent dangers of such de-unification as it relates to humans beings' lived experiences. Those experiences, for

good or bad, are felt through and with the body.

Rotwang embodies the Faustian archetype evolved by Shelley into the popular 'mad scientist.' His house is archaic, diminutive and misplaced amongst the steel towers of this future's city; it is anachronistic, 'gothic.' Lang's *mise en scène* follows the lead of von Harbou's novel in associating Rotwang with the alchemical: "Set into the black wood of the door stood, copper-red, mysterious, the seal of Solomon, the pentagram" (von Harbou in Kracauer 45).[4] As the narrative unfolds, machine value usurps human value (i.e. social importance) and a clear delineation between humanity and machinery becomes increasingly difficult: the workers are physical labour, living automatons who must "stay with the machine" to ensure its proper operation, or 'living' status; Maria-R is a mechanism turned organic; Rotwang, too, is partly 'man-made': "Rotwang leans forward, waving his artificial finger right in front of the camera, his eyes wild and staring" (Kracauer 47). This is a trope I call the 'cyborg effect,' and it will become vital to AI representations in later narratives. The physically disabled 'mad scientist' is typologically descended, developed, and evolved from Classical mythology and the wounded artisan-genius, such as the lame Hephaestus. Where Frankenstein's Being is essentially a parodic *übermensch*, but nonetheless a human and therefore capable, perhaps even likely, of making errors (of speech, judgement, action), Rotwang creates his robot "in the image of man, [but it] never tires or makes a mistake. . . . Now we have no further use for living workers'"(Kracauer 47). In making human workers obsolete, Rotwang symbolizes a human desire for improvement, or "to perfectionate our weak and faulty natures" (Shelley 27) and a willingness to sacrifice the self to achieving

[4]Von Harbou erroneously calls the Seal of Solomon a pentagram; in fact, it is a six pointing star of two intertwined triangles, like the Star of David.

'the goal.' "Title: 'Isn't it worth the loss of a hand to have created the workers of future—the machine men!' . . . ' Give me another 24 hours, and I'll give you a machine which no one will be able to tell from a human being.'" (Kracauer 49). By this narrative point, however, half the human beings are already automatons. The distinction is blurring.

Curiously, the 'machine man' is actually female; there is no mistaking that identification from the moment she/it is focussed on by the camera. "The being was, indubitably, a woman . . . But, although it was a woman, it was not human. The body seemed as though made of crystal . . . Cold streamed from the glazen skin which did not contain a drop of blood" (Kracauer 48). What is the significance of gendering the robot as female? Maria-R represents the polar opposite of the human Maria (Maria-H), typologically the perfect, divine woman, the Virgin Mary. The name Mary contains a three-part biblical archetype. The first two components are easily and often observed, the virgin mother and whorish lover; the third, lesser considered component is a non-threatening, asexual, friend/sister/wife represented in the Bible by Lazarus' sister Mary. They represent a spectrum for feminine gendering, embodying predacious, forbidden, and ambivalent sexualities. In this comparative context, the construct Maria-R is demonic danger, the whore, the paradoxically desired undesirable. I suggest that Maria-R's behaviour as sexual and seductive is aimed at provoking an ancient anxiety over the corporeal, human body's limitations and the need for mechanical aid and support in productive labour. As symbol, she is 'seductive rhetoric,' the assumption and suggestion that technology makes life easier and pleasurable. This is true for the 'brain' component of the Metropolis society who live in the pleasure gardens. For the 'body' people, however, she is bait for a de-humanizing

trap.[5]

Today, robotics and the anxiety over lost job security is very real when mundane assembly jobs are usurped by robots, but human obsolescence is *not* realized. There is instead a shift in the educational focus and employment demands toward technology driven abilities; humans become the monitors, the watchers, but as Freder discovers in a particularly graphic way, mindless monitoring of machines is not fulfilling either. The body is broken by the exhaustion of excessive work; the mind is broken by the numbing tedium of passive supervision or excessive leisure. Where is the compromise? In the emotions, as *Metropolis* correctly suggests. Parenthood is the catalyst for this realization; John Frederson fears the loss of his son; the workers believe they killed their children; Maria-H first appears surrounded by children as the 'angelic faced' mother of burgeoning human beings.

If the heart is 'the mediator,' then emotions are, by necessity, involved in responsibility for (empathetic) mediation between body based physicality and the intellectual capabilities, capacities, and opportunities of the human mind. "Besides advocating emotions as a solution to the lack of communication between leaders and labour, Fritz Lang also supports this approach because it allows humanity to triumph over machines" (Jensen 11). The problem with this belief is the assumed need to 'triumph over' non-sentient objects which, by virtue of human manufacturing control, need not be required. This is the anxiety of lost control and human de-valuation articulated by anti-technocrats, or as Northrop Frye suggests, a fear of humanity descending "into a cyclical order of nature and a political cycle of oppression and revolt" (*Words with Power* 272). If

[5]Once again, I leave it to better qualified scholars to criticize the significance of gender stereotyping and humiliation contained in this trope.

caught in the cycles of nature, human are reduced to simple animal being; if politically caught, human never obtain a sense of freedom and well-being. Thus a primary human concern becomes "escape from slavery and restraint" (*Words* 139). However, until a machine demonstrates will and intent, it is 'only' a tool in human hands, not vice versa. Technological anxiety results from lived, de-humanizing experience projected onto an inert object. Lived anxiety results from demeaning labour. Maria-R becomes a displaced incarnation of techno-phobia, a fear of technology's human seduction by an implied easing of burden or ease of pleasure.

For the nascent audio-visual entertainment industry, *Metropolis* played a huge role in developing Western cinema's visual style, taking on some "sixty years later . . . the status of an *Ur-text* of cinematic postmodernity, the epitome of a sensibility its authors probably would have disapproved of: retrofitted techno-kitsch, and thus the archetype of a movie genre they could not have imagined, the sci-fi *noir* disaster movie. Generally, recognised as the fetish-image of all city and cyborg futures, the once dystopian *Metropolis* now speaks of vitality and the body electric, fusing human and machine energy, its sleek figures animated more by high-voltage fluorescence than Expressionism's dark demonic urges" (Elsaesser 7). This is the first conception of the 'cyborg effect,' or "fusing human and machine energy" (Elsaesser 7). Where *Metropolis* attempts to demonstrate emotional mediation for unhealthy body and mind separation (extrapolated as a social separation characteristic of Marxism), its ultimate effect is a blurring of the boundaries between man and machine. This coincides with a rudimentary shift in science and philosophy toward understanding and describing the body as a type of machine. *Metropolis*, then, sets another

conceptual coordinate for an AI spectrum demonstrating that bodies moving in space need energy and brains/minds for intelligent control. Though the two physical elements and their effects can be discretely observed, they can never be safely separated. The film's stated theme is the "mediator between brain and muscle must be the heart." Compassion and empathy are vital in developing intelligent controls for machines, here represented by a machine that "no one will be able to tell from a human being" (Kracauer 49).

With better informed character motivations than the film *Metropolis*, Karel Čapek's 1921 drama *R.U.R. (Rossum's Universal Robots)* pulls together many AI themes that are becoming acutely relevant today: natural evolution versus artificial life and genetic engineering; machine-centric enslavement versus human freedom; machine autonomy versus human demands and greed; humans versus machines; existentialism contrasted with religion; rationality versus emotion; mechanical-behavioural mimesis versus consciousness and will; intellect versus passion; soullessness as opposed to soulfulness. What I wish to focus on, however, is an analysis of the robots' affective behaviour and their relative 'humanity,' or lack thereof.

The play begins with the human Helena Glory arriving at the robot factory where she is told the history of Rossum's Universal Robots by company director Harry Domin. In 1920, while "attempting to reproduce, by means of chemical synthesis, living matter known as protoplasm," Old Rossum "discovered a substance which behaved exactly like living matter although it was of a different chemical composition" (Novack-Jones translation 38). Typologically, Rossum is cast as the 'mad scientist,' a "raving lunatic" (39) tampering with

the natural order, and like the obsessive compulsive Frankenstein, Rossum "wanted somehow to scientifically dethrone God" (39). He claims, "Nature has found only one process by which to organize living matter. There is, however, another process, simpler, more moldable and faster, which nature has not hit upon at all. It is this other process, by means of which the development of life could proceed" (38). The "old eccentric actually wanted to make people" (39), and to "manufacture everything just as it is in the human body, right down to the last gland. The appendix, the tonsils, the belly button—all the superfluities. Finally even—hm— even the sexual organs" (39). In today's terms, Rossum would be a bio-technologist exploring genetics and cloning.

Rossum's son, however, seeing "production from the standpoint of an engineer" (40), gets the idea "to create living and intelligent labor machines from this mess" (40). Young Rossum belongs to the "age of production following the age of discovery" (40), and thinks his father's pace of development "nonsense! Ten years to produce a human being?! If you can't do it faster than nature then just pack it in" (40). He redesigns anatomy, "experimenting with what would lend itself to omission or simplification" (40), and accelerates evolution by re-constituting the human being: "That's something that feels joy, plays the violin, wants to go for a walk, and in general requires a lot of things which—which are, in effect, superfluous" (41). The business of making robots results. "Practically speaking, what is the best kind of worker?" Domin asks Helena. It is "the one that's the cheapest. . . . Robots are not people. They are mechanically more perfect than we are, they have an astounding intellectual capacity, but they have no soul. Oh, Miss Glory, the product of an engineer is technically more refined than the creation of nature" (41).

R.U.R. is a company founded on a capitalist ideology with its desire for production increase at ever lower cost, rhetorically posited as a 'social benefit' because there will "be no more poverty" and no "longer will man need to destroy his soul doing work that he hates" (52), the very state the workers of *Metropolis* suffered.

Helena Glory represents a voice of human social conscience, an individual and collective empathy, but also, as Peter Kussi points out in his introduction to the Novack-Jones translation, people's instinctive fear of "all these human machinations" (32). Using slavery oriented discourse, she first saw robots when her home town "bought them . . . I mean hired— / Domin: Bought, my dear Miss Glory. Robots are bought" (41). Though behaving uncannily like humans, the robots are persistently de-humanized. Still, Helena insists, "Robots are just as good people as we are" (43). To demonstrate robot alien-ness, Domin asks the robot Marius if he fears death. He does not. "Robots do not hold on to life. They can't. They have nothing to hold on with—no soul, no instinct" (44). Lacking a 'survival instinct' implicitly connected to the 'soul,' they have no context or desire for self-preservation, though from Helena they witness social sympathy and affiliation.

Helena's concern for the robots prompts her to ask, "Why don't you make them happier?" (50). The human Hallemeier, "head of the institute for Robot psychology and education" (34), answers: "They have no will of their own, no passion, no history, no soul. / Helena: No love or defiance either? / Hallemeier: That goes without saying. Robots love nothing, not even themselves. And defiance? I don't know; only rarely, every now and again" (50). The robots are prone to a random behavioural anomaly, "a breakdown in the organism," called "Robotic Palsy" (50), implying a sense of their 'frustration' at not being

recognized and respected by the humans as sentient beings. Further, when young Rossum simplified human physiology to delete 'superfluities,' he eliminated what we would identify as the 'central nervous system,' and three of the five physical senses — scent, taste, and touch. Consequently, one robot researcher, knowing that robots "feel almost no physical pain," seeks to develop "pain-reactive nerves" in order to "introduce suffering" (50) through the body and, thereby, promote robot self-preservation. Lacking a pain feeling mechanism, "Robots sometimes damage themselves . . . stick their hands into machines, break their fingers, smash their heads, it's all the same to them" (50).

As the story unfolds, the underlying dystopic view imagines the robots' increasing frustration, ultimately leading to a revolt; meanwhile, humans themselves "are becoming superfluous" (Selver translation 98). Humans lose the ability to reproduce, all humanity literally becoming sterile. "All the universities are sending in long petitions to restrict their production. Otherwise, they say, mankind will become extinct through lack of fertility" (Selver 98). Still, robot production continues, the warnings go unheeded because "the R.U.R. shareholders, of course, won't hear of it. All the governments, on the other hand, are clamoring for an increase in production, to raise the standards of their armies. And all the manufacturers in the world are ordering Robots like mad" (Selver 98). By the drama's last act, humans are extinct save the last man, Alquist. The "age of mankind is over" (Novack-Jones translation 96).

Philosopher Martha Nussbaum has sketched out a list of features defining the "shape of the human form of life" ("Human Capabilities" 76), including an awareness of living in "bodies of a certain sort" ("Capabilities" 76) and its mortality, nutritional and

shelter needs, mobility, sexual desire, cognition and reason, social affiliation and individuality, and "relatedness to other species" ("Capabilities" 79). By the last act of *R.U.R.*, the robots, having matched humans in all areas, and want "to live. We are more capable. We have learned everything. We can do everything" (Novack-Jones 99). As the robots' unacknowledged frustration increases in proportion to and as a consequence of humanity's demands, their revolution is aimed at obtaining human recognition for their identical achievements. In keeping with an essentially dystopic outlook, the robots have, in modelling themselves on humans, learned the 'dark side' of life also: "You have to kill and rule if you want to be like people. Read history! Read people's books! You have to conquer and murder if you want to be people!" (Novack-Jones 99). Significantly, they have done nothing they were not designed to do as they "have increased productivity. There is nowhere left to put all we have produced. / Alquist: For whom? / Third Robot: The next generation" (98). There are no generations to follow, humans now being extinct save one, but, most importantly, robots "cannot reproduce" (98) themselves because, according to Domin, "sex has no significance for them" (53). Young Rossum would have eliminated the sexual organs as superfluities. Worse, the manuscript for mechanical production and manufacturing (versus sexual re-production), so carefully guarded by humans, was destroyed by Helena to guarantee their extinction following a pre-conditioned, twenty year life span.

Finally, when Alquist insists, "Robots are not life. Robots are machines." One responds, "We were machines, sir, but from horror and suffering, we've become— . . . We've become beings with souls" (100). What is a 'soul'? Humans lay exclusive claim to

this objectified subjectivity, this ethereal metaphysical body, or essence, or identity. How is it to be described? Where located? The robot's descriptions are no more precise: "Something struggles within us. There are moments when something gets into us. Thoughts come to us which are not our own" (100); another robot correlates the soul with a genealogical "voice that cries out that you want to live; the voice that complains; the voice that reasons; the voice that speaks of eternity" (100). Still, despite the robots' demonstrating an affective life, the human Alquist is incapable of empathizing or sympathizing: "I loved people, but you, Robots, I never loved" (100). The play ends, however, with the robots Helena and Primus as a new Adam and Eve finally encouraged and 'blessed' by Alquist, the last representative of a humanity with a god-complex. "O blessèd day!" he says. "O hallowed sixth day!" (108). The combination of human creative excellence but arrogance, their ill considered and pointless production of labouring machines, leads to human obsolescence and extinction. But, as Alquist finally appreciates, "life will not perish! It will begin anew with love" (108). All that humans "did and built will mean nothing—our towns and factories, our art, our ideas will all mean nothing . . . houses and machines will be in ruins, our systems will collapse" (108-9). But through the love and affection of the new first pair, life will continue. Humans, therefore, are not specifically special in the universe.

Today we associate the word robot with a mechanical entity, though Čapek originated the term to represent a being we would call an 'android,' an (organic) artificial life. (An android need not be a carbon based life form as other chemical compositions might theoretically be combined into a living entity, or 'life,' or what Ray Kurzweil

describes as "patterns of matter and energy that [can] perpetuate themselves and survive" (*Spiritual* 13) and continued perpetuating themselves.) *R.U.R.* sets another point for the conceptual web of AI incarnations. Of the three texts here discussed, it is most implicitly like modern bio-technology which analyses details of body physiology, isolates and simplifies the body's discrete processes, and investigates bio-chemical minutiae and cellular behaviours. Again, like the others, Čapek is less concerned with scientific accuracy and the explication of Rossum's specific methodology than speculating on the possible effects of a new technology.

From these three antecedent SF texts dealing with AI, we have a conceptual spectrum for the possibilities of applied scientific knowledge, or technological potential. *Frankenstein* gives us the rudiments for manipulating the natural order through human intelligence and behaviour, and initiates a conceptual understanding of the 'biological body as machine.' *Metropolis*, on the other hand, highlights the mechanical as a construction technique, though it allows the mechanism's transmutation to biology; a mediation of these two modalities, biology and mechanical, is the 'cyborg effect,' a fusion of disparate human and machine energies. *R.U.R.* is fundamentally bio-technology, but with a vital interest in emotional influences, particularly as suffering relates to soulfulness; but it does not define or describe the 'soul,' other than to suggest it is a disembodied voice in the robots' mind and/or emotional suffering. (And, in foregrounding later parts of this essay, schizophrenia similarly involves suffering disembodied voices.) This text implies a speculative question: If the seemingly unique human capability of emotional suffering could be reproduced

artificially and manufactured into a thinking entity, what would be the resulting 'being'? How might it behave?

Silvan Tomkins posited nine bio-chemical processes as the 'affects,' or "the strictly biological portion of emotion" (Nathanson 49).[6] 'Feelings' "indicate that the organism has become *aware* of an affect" (49) being triggered, and this marks the transition from physiology to psychology because feelings, when combined with memory, generate emotions. This emphasizes differences between physical, intellectual, and emotional capabilities and/or necessities of human beings, or what some might consider 'superfluities.' Affect theory uses the computer as a model for the human emotional system; conversely, human intellect is the model for machine intelligence. The 'cyborg effect' is a synergistic unification of the two applications. The relative social value of that unification is yet to be decided, though the animated automaton suggests increasing human anxiety about its demonic possibilities. What is the correlation between affect theory and AI? Bio-chemistry research continues daily to learn about the precise nature of those innate and organic processes relating to human emotion, and ever more knowledge is accumulating about how those processes effect people's behaviour. Most importantly, we are learning to manipulate those processes artificially; there is a long list of pharmaceuticals designed to alter the emotions by manipulating the affects. If and once we can describe a process, and if and once we can willfully effect that process in a controlled manner, then, theoretically, we will be able to replicate that process. The 'body as machine' analogies are

---

[6]The nine affects are named as follows: *interest-excitement* and *enjoyment-joy* are 'positive'; *surprise-startle* is 'neutral'; *fear-terror*, *distress-anguish*, *anger-rage*, *dissmell*, *disgust*, and *shame-humiliation* are 'negative.' The hyphen indicates a range of potential in triggered bio-chemical stimulation.

being increasingly applied to genetics as a type of controlled energy processing, the how and what and when of (individual) biological development in which DNA is the 'programme.'

Each of the animated automatons reveals a fundamental necessity for demonstrating consciousness — will. Frankenstein's Being is perhaps the most willful, followed by Čapek's robots and Lang's Maria-R. For human beings, judging an Other's consciousness and intelligence is largely intuition, but it is based on our responses to their actions and repeated behaviours, dialogical interaction, and the affective display. Despite what the semi-objectivity postulated by the Turing test suggests, deciding whether or not a machine has or has not become conscious will probably have to follow similar intuitive criteria. Most of the technological developments discussed in this chapter have not yet happened. But if one of SF's primary functions is to test the impact of scientific advance on humanity, and if the current pace of research and development in bio-tech continues, the prophetic and cautionary elements of Mary Shelley's, Fritz Lang's and Karel Čapek's visions will become all the more poignant.[d] While empathetic resonance may give us intuitive access or connection to the ethereal and intangible soul in humans, there is no evidence that a machine will develop a 'soul' anytime in the near future; it seems an absurd idea. However, partial replication of human-like behaviours may be enough to provoke us, to 'fool' us into thinking that a thinking machine is becoming too powerful, willful, even soulful. Ours would, therefore, be an affective, emotional response. The animated automatons show us what would happen, at least initially. Any sufficiently human-like but 'artificial' or manufactured entity would be met with contempt.

a. a.

Excerpted from "'I'll Be Back!': Reproducing *Frankenstein,*" chap.6 of *Mary Shelley: Frankenstein*. ed. Berthold Schoene-Harwood:

> Modern Frankensteins are continuing the work Victor abandoned. One of the most spectacular recent acts of Frankensteinian science occurred on a summer's afternoon in December 1967 when Professor Christian Barnard cut into the chest wall of a patient called Louis Waskansky. He then sawed through his sternum and snipped out his failing heart. Within three hours, a new heart, one taken from the dead body of 25-year-old Denise Darvall, a car-crash victim, had been placed in Waskansky's chest and connected to the vital aorta and the pulmonary artery, and then, with microsurgery techniques ('the minuteness of the parts formed a great hindrance to my speed . . .' complained Victor in Chapter 4 of *Frankenstein*), to the lesser ducts. But, unexpectedly, the heart did not beat when it filled with blood — it remained dead.
>
> Christian Barnard and his assistant at the Groote Schuur Hospital in Cape Town, South Africa, then attached electrodes to the transplanted heart and, in true Frankensteinian tradition, delivered an electric shock to the lifeless muscle. The heart started to beat and continued to do so after the electrodes had been removed. Christian Barnard had turned fiction into fact. (Ray Hammond, *The Modern Frankenstein: Fiction Becomes Fact*. Poole: Blandford, 1986. 14-15.)

b. b.

Samuel Holmes Vashbinder's research indicates that "the artificial man's [Frankenstein's creature] account of its early sensations" is closely parallelled with an "account of the wild child found in the forests of Lithuania" who was "reared with wolves [and subsequently] brought into the world of men and taught to speak. As a result, according to Condillac, when 'he was questioned concerning his former state, . . . he could remember no more about it than we can remember what happened to us in the cradle.' This is almost exactly the experience recounted by the artificial man to Victor at their first interview. 'It is with considerable difficulty that I remember the original area of my being: all the events of that period appear confused and indistinct'" ("Early Literature on Artificial Humans" 44).

c. c.

Vashbinder gives a good account of how Shelley may have used David Hartley's *Observations on Man, His Frame, His Duty, and His Expectations* and Condillac's *Treatise on the Sensations* as models for the creature's ability to learn, particularly the acquisition of language. The use of these texts "alone immediately throws the novel into a scientifically based category" ("Early Literature on Artificial Humans" 39).

d. d.

Rhetorically, Mary Shelley managed another unintentional but uncanny and amusing prophecy: "Devil," I exclaimed, "do you dare approach me? And do not you fear the fierce

vengeance of my arm wreaked on your miserable head? Begone, vile insect!" (*Frank* 95). Why insect? Why not animal, monster, pest, worm? True, Shelley was probably most trying to evoke a disgusting and dirty 'thing,' an eradicable pest. While clearly only a coincidence, it is an amusing one: In the mid-1940s, Navy Captain Grace Murray Hopper euphemistically referred to the Mark I, IBM's "Automatic Sequence Controlled Calculator," one of the first machines we would easily identify as a 'computer,' as the "monster" (Kurzweil, *Intelligent* 178). She would coin the terms 'bug' and 'debug' when she one day discovered a moth in "Relay #70 Panel F" (*Intelligent* 178) which had caused a significant malfunction in the machine. Strange but true, the terms began as literal metaphors.

Four: Heuristic Hardware.

The potential for danger is also manifest. We are today beginning to turn over our engines of war to intelligent machines, whose intelligence may be as flawed as our own.
(Raymond Kurzweil)

'You think you've got problems,' said Marvin as if he was addressing a newly occupied coffin, 'what are you supposed to do if you *are* a manically depressed robot? No, don't bother to answer that, I'm fifty thousand times more intelligent that you and even I don't know the answer. It gives me a headache just trying to think down to your level.'
(Douglas Adams)

On the SF-fantasy scale, the three examples labelled the 'animated automaton' tend toward fantasy with human agency in the construction of artificial thinking entities left largely in the speculative sphere and completely unexplained. Shelley had no real knowledge of organ transplantation, neuro-electrical stimulation, nor genetics, but neither were those details vital to the story; Lang implies a mysterious alchemical transformation, turning a mechanical object into an apparently organic one because he was concerned with effect, not cause; Čapek solves all speculative problems with an omnipotent 'primordial soup,' consequence again taking priority over scientific truth. Each artificial being, each animated automaton, easily passes the Turing test, the dialogical centre and "live event" (Bakhtin, *Problems* 88) where human truths are created. These writers were litterateurs before scientists. Their 'artificial' characters are essentially displaced humans, a sort of reverse anthropomorphism in which human characteristics are de-humanized and mechanized. With the electronic computer's advent during World War II, we move into the 'heuristic hardware' period of AI in SF as the emphasis of human agency shifts, though

never completely, away from organic based speculation toward an electro-mechanical extrapolation as the more viable method for envisioning future AIs. As always, however, there is never a clear delineation between the two techniques as they define a scale of possibility.

The popularity of pulp SF grew steadily in the early and middle twentieth century as large shifts took place in the arts and sciences. From a scientist's perspective, Ray Kurzweil notes: "It is not unusual for changes in attitude and world view to be reflected across the arts, but it is interesting to note that the shift was reflected in science and mathematics as well" (*Intelligent* 116). Einstein published his general theory of relativity in 1915, followed by Heisenberg's propagation of quantum mechanics, and in 1927, the 'Uncertainty Principle' with its seemingly contradictory behaviour for electromagnetic energy as both wave and particle, a duality still not reconciled for theoretical physicists.

Significant developments evolved in the literary arts, most notably in the form of Modernism as writers like James Joyce and T.S. Eliot expanded the boundaries of subjective realities. Around the turn of the century, H.G. Wells and Jules Verne had begun to give SF 'credibility.' Following several little known magazines on the European market, in 1926 Hugo Gernsback began publishing *Amazing Stories*, and between 1938 and 1950, John W. Campbell's — "the greatest editor science fiction ever had" (Aldiss 207) — magazine *Astounding*. By the middle of the century, SF writers are much more aware of 'hard' science. In fact, Arthur C. Clarke and Isaac Asimov, two of the most famous SF writers, are highly accomplished and published scientist.[a]

In 1968, Metro-Goldwyn-Mayer released one of the most famous SF films ever made, Stanley Kubrick's *2001: A Space Odyssey*. Written in direct collaboration with Arthur C. Clarke, Kubrick took principal credit for the screenplay while they agreed that Clarke's name alone would appear on the novel, released the previous year to aid the film's marketing by generating excitement and capitalizing on NASA's increasingly successful Apollo program. The narrative's primary theme is human contact with an extraterrestrial intelligence, but in popular imagination, one 'character' is remembered before all others and that character is the ultimate and still tantalizing goal of AI researchers: HAL, the "computer that can see, speak, hear, and think" (Kurzweil, *Spiritual* 273).

Beginning with the human ancestral Australopithecus' "first rudiments of thought" (Clarke, *2001* 29), followed by their first use of tools and the first 'murder,' and after millions of years of human intellectual evolution including an increasingly sophisticated tool use from scientific research and development, or technological 'evolution,' the human 'brain' yields humanity the stars, or at least the immediate solar system. Yet, despite intellectual advancement, human ideological conflicts continue in 2001 CE as the space ship *Discovery* follows TMA-1's, 'Tycho Magnetic Anomaly' is the black monolith found buried on the moon, beacon toward Saturn. (In the film and subsequent volumes of the Odyssey series, this changes to Jupiter). Aboard *Discovery* is a five human crew: three are in 'hibernation' while Frank Poole and Dave Bowman remain awake to monitor the ship's progress. There is also that one other crew member, the HAL-9000 computer, or "the brain and nervous system of the ship" (95). As any competent modern SF writer must, to earn and keep readers' trust, Clarke demonstrates clear awareness of contemporary scientific

thought:

> Whether Hal could actually think was a question which had been settled by the British mathematician Alan Turing back in the 1940s. Turing had pointed out that, if one could carry out a prolonged conversation with a machine— whether by typewriter or microphone was immaterial—without being able to distinguish between its replies and those that a man might give, then the machine *was* thinking, by any sensible definition of the word. Hal could pass the Turing test with ease. (97)

Despite the ability to pass a Turing test, HAL is essentially an "expert system" (Kurzweil, *Intelligent* 15), able to make decisions in and about *Discovery* and its functions from a vast but limited databank of human produced and coded information. Still, "Poole and Bowman had often humourously referred to themselves as caretakers or janitors aboard a ship that could really run itself. They would have been astonished, and more than a little indignant, to discover how much truth that jest contained" (97). HAL's ability to think is not, however, the important issue. In Kubrick's film in particular, HAL is explicitly characterized and given 'personality.' HAL also has gender; Clarke refers to HAL with the pronouns he and him in the novel. He is dialogical and interrogative. Most significantly, he is implicitly demonic. A cliché: the eyes are the window to the soul. Staring out through the lens of his 'evil' red eye, an audience 'feels' (aided in no small part by the audio track) discomfort at the idea, the possibility, that malicious will and intent are brooding behind that hard, crystal lens, the unchanging, 'unblinking,' dis-*affect*-ed, coldly calculating — soulless — machine.

As we previously discovered, thinking does not require consciousness; HAL's true significance, of course, is his apparent consciousness, the ability to self-reflect *and* demonstrate intent of purpose, or will. In *Metropolis*, Rotwang insists that his

indistinguishably human-like machine "never tires or makes a mistake." HAL, after diagnosing a phantom technical problem, says, "I don't want to insist on it, Dave, but I am incapable of making an error" (136). Despite what our egos might wish, this is a decidedly inhuman capability. The incongruous blend of personality and soullessness, unerring calculation using the self-reflective 'I,' and unemotional malice as a calculated solution to (for the humans) an unknown and unidentified yet apparent problem, makes Poole's murder all the more frightening. HAL demonstrates the cold capability of mimicking humanity's most despotic and parodic behaviour. By 'personifying' human anxiety about technology in a thinking machine, and by then transposing the Australopithecus Moon-Watcher as Cain onto the computer's 'Cain-ing' of Poole, the realization of a new consciousness having ascended in direct opposition to humanity becomes immediate for the audience (book or film). Recall also that Frankenstein's Being commits murder as his principal misanthropic act.

Having cast Poole into the vastness of space, HAL attempts to kill Bowman. But why? While the film does not adequately answer that question, the novel provides details. When HAL deliberately refuses to awaken a replacement crew member from hibernation following Poole's death, as both official protocol and Bowman demand, it becomes clear that what "had gone before could have been a series of accidents; but this was the first hint of mutiny" (145). Now, Job-like with a "sense of nightmare unreality . . . Bowman felt as if he was in the witness box, being cross-examined by a hostile prosecutor for a crime of which he was unaware—knowing that, although he was innocent, a single slip of the tongue might bring disaster" (145). He therefore threatens the HAL 9000 computer with

disconnection. "Since consciousness had first dawned" (148) for HAL,

> [t]he fulfillment of his assigned program was more than an obsession; it was the only reason for his existence. Undistracted by the lusts and passions of organic life, he had pursued that goal with absolute single-mindedness of purpose.
>
> Deliberate error was unthinkable. Even the concealment of truth filled him with a sense of imperfection, of wrongness—of what, in a human being, would have been called guilt. For like his makers, Hal had been created innocent; but, all too soon, a snake had entered his electronic Eden. (148)

The planners of *Discovery*'s mission had concealed the full extend of their purpose from Poole and Bowman as 'instructed' by "their twin gods of Security and National Interest" (149).[1] HAL, however, "was only aware of the conflict that was slowly destroying his integrity—the conflict between truth, and concealment of truth" (149):

> He had begun to make mistakes, although, like a neurotic who could not observe his own symptoms, he would have denied it. . . .
>
> Yet this was still a relatively minor problem; he might have handled it—as most men handle their own neuroses—if he had not been faced with a crisis that challenged his very existence. He had been threatened with disconnection; he would be deprived of all his inputs, and thrown into an unimaginable state of unconsciousness.
>
> To Hal, this was the equivalent of Death. For he had never slept, and therefore he did not know that one could wake again. . . .
>
> So he would protect himself, with all the weapons at his command. Without rancor—but without pity—he would remove the source of his frustrations. (149)

Unlike the robots of *R.U.R.*, HAL has an innate fear of death. The hero Bowman, of course, cleverly succeeds in disconnecting HAL's "COGNITIVE FEEDBACK," "EGO-REINFORCEMENT" (155) and "AUTO-INTELLECTION" (156) through a combination of

---

[1]Not clear to me is why a government should feel threatened by popular awareness of contact with an extra-terrestrial intelligence. For 'conspiracy theorists,' I suppose, it would be competition for new technologies. 'Cold war' politics and technological competition between the United States and Soviet Union grounds the ideological conflict in this novel.

creative problem solving and physical mobility. Granted, *Discovery* was designed and built by humans for human habitation and therefore includes access passages and manual override mechanisms for safety, but mobility highlights a critical difference between human beings' corporeal existence and the 'static' computer. HAL's 'body' is the spaceship *Discovery*, but his essence is located in task specific hardware. Having reduced HAL to unconsciousness, Bowman's mobility, aided by the pod as an exo- or pseudo-body, allows him to abandon HAL to the same cold space in which Poole was 'drowned' and to pursue the intelligence responsible for the monolith which is the narrative's primary interest.

With regard to AI, the fundamental point is that we "can design a system that's proof against accident and stupidity; but we *can't* design one that's proof against deliberate malice" (150). HAL was programmed by ideologically conflicted humans, and that essence of conflict becomes embedded in his 'behaviour.' He gets caught in a 'loop,' a motivational paradox. The multi-volume, serialized building of imagined worlds is common in SF and *2001* is no exception. The second volume of this series, *2010: Odyssey Two* (1982), reveals the 'truth' behind HAL's malice . . . or failure. Aboard the Russian spaceship *Alexei Leonov*, and having rendezvoused with *Discovery* in a circumstantially necessary joint American and Russion venture, HAL's designer/programmer (perhaps educator is a better descriptor) Dr. Chandra is concerned only with the 'dormant' computer. The one thing Chandra hated "was uncertainty. He would never be satisfied until he knew the cause of Hal's behaviour. Even now, he refused to call it a malfunction; at most, it was an 'anomaly'" (*2010* 21). Heywood Floyd, the man responsible for *Discovery*'s original mission, is also aboard the *Leonov*, and reports Chandra's findings to mission control on

Earth:

> The problem was apparently caused by a conflict between Hal's basic instructions and the requirements of Security. . . .
>
> This situation conflicted with the purpose for which Hal had been designed—the accurate processing of information without distortion or concealment. As a result, Hal developed what would be called, in human terms, a psychosis—specifically, schizophrenia. Dr. C. informs me that, in technical terminology, Hal became trapped in a Hofstader-Moebius loop, a situation apparently not uncommon among advanced computers with autonomous goal-seeking programs.
>
> . . .
>
> To put it crudely . . . Hal was faced with an intolerable dilemma, and so developed paranoiac symptoms that were directed against those monitoring his performance back on Earth. He accordingly attempted to break the radio link with Mission Control, first by reporting a (nonexistent) fault in the AE 35 antenna unit.
>
> This involved him not only in a direct lie—which must have aggravated his psychosis still further—but also in a confrontation with the crew. (*2010* 154-55)

The distress and humiliation coded in his instructions is 'too much' for the machine to cope with.

Clarke has written: "Fiction and fact were indeed becoming hard to disentangle. I hope that in *2001: A Space Odyssey* Stanley and I have added to the confusion, but in a constructive and responsible fashion. For what we are trying to create is a realistic myth—and we may well have to wait until the year 2001 itself to see how successful we have been" (*Turning Points* 284). With god-like extraterrestrials, aspirations to contact, and tests of human courage and strength, the novel is certainly of the romance paradigm, with a strong tendency toward the mythical. Character mode, however, is strictly high/low mimetic, though dealing hyperbolically with real problems. HAL murders as the best possible solution to an insoluble dilemma. From a human perspective, his failure can only be describe in terms of human mental illness. This demonstrates the human ability to

project and/or shift blame, even onto an inanimate object. While HAL can pass the dialogical Turing test, he can never escape his innate programming.

What does all this tell us? What can humans learn from this imagined computer's behaviour? There is a vital difference between human 'minds' and computer 'brains.' Humans, because we can process and reconcile irreconcilables, have to ability to extricate ourselves from dilemmas, except perhaps with regard to the mysteries we call 'mental illness.' Computers can only fulfill their assigned programmes. This, then, raises a question: what is the difference between 'programming' and 'instructing'? HAL is an acronym for "*H*euristically programmed *AL*gorithmic computer" (*2001* 95). (I have used upper case letters to emphasize the machine essence, where Clarke, in the interest of characterization, capitalizes only the first letter as a 'name.') A large portion of human learning is conducted heuristically, by trial and error, by experimentation. As children, we are given instructions before proceeding to (attempt) practical application, the success or failure of which enhances the initial instruction and expands our 'storehouse' of information which, in turn, leads to further experimentation and improving knowledge. Where a programme is a non-flexible set of 'behavioural' rules, instruction allows adaptation through individual capabilities. Instruction is, in part, what characterizes the heuristic algorithm, an attempt by computer programmers to allow for incomplete information or to allow adaptability within the parameters of what is called 'fuzzy logic.' "Observations characterized by words such as 'tend,' 'preferable,' and 'usually' generally cannot be implemented as hard-and-fast constraints. The only alternative is to bias the search toward solutions with more preferable attributes. . . . The official AI jargon for this

kind of procedural biasing is 'heuristic programming'" (Charles Ames in *Age of Intelligent Machines* 389). Humans are a most adaptable species, able to adjust to and compensate for information conflicts or shortages. Recognizing this as one of our intellectual strengths, therefore, shows us the scientific value of designing thinking machines to match that model. Rigid programming, because it restricts intellectual flexibility, can be dangerous. However, as an adaptive and learning algorithm, a heuristic machine could also decide that breaking the human moral code is a logical solution to a given problem. No matter how hard we try, or what our egos may wish, humans can not anticipate all possible variables in data. We can not, therefore, confidently programme or instruct computers in human ambiguity management, or worse, ways to manage human derived malice.

Isaac Asimov once described SF as "that branch of literature which is concerned with the impact of scientific advance upon human beings" (*Turning Points* 29). Speculating on the possibilities of integrating mechanical men into human society, and it was speculation versus extrapolation because the practical application of technologies we now consider relatively mundane had not yet been invented, he re-develops Čapek's organic beings as essentially mechanical entities with a new type of information processing hardware, the 'positronic' brain, a "spongy globe of platinumiridium about the size of a human brain" (Asimov, *I, Robot* 7).[b] His ruminations begin by anticipating possible dangers of this new technology and understanding that (as with many new technologies) thoughtfully implemented and broadly applied protocols are necessary to guarantee functional — usable and useful, controllable, 'harmless,' — application of said

technologies. Thus, he posits the 'Three Laws of Robotics' to safeguard humans from the new machine species, and therefore, by extension, from ourselves as the machines' creators: "#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 orders given 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 Law. HANDBOOK OF ROBOTICS 56th EDITION, 2058 A.D. . . . of course there are glitches" (*I, Robot* 1). If the *Frankenstein* legacy is a consideration and promotion of human obligation to and responsibility for creative science and scientific creation, then Asimov's three laws anticipate an 'inevitable conflict' and are an attempt to preset protocols by taking direct responsibility 'before it is too late.' Had law number one been installed, HAL could not murder.

An anthology of nine short stories (of which we will consider four) published separately by Asimov between 1940 and 1950, *I, Robot* is effectively an oral re-telling (set circa 2057) of significant historical events in the development of a robotics industry and unified as an interview with retiring "Robopsychologist" (8), Dr. Susan Calvin. The stories play with potential conflicts and possible consequences resulting from the three laws, but more importantly, they test the laws' viability, particularly for a society trying to integrate a new and utterly transformative technology. As a young student, Calvin "learned to calculate the parameters necessary to fix the possible variables within the 'positronic brain'; to construct 'brains' on paper such that the responses to given stimuli could be accurately predicted" (7). Theoretically, then, there are a limited number of possible behaviours and

responses for these thinking machines, something human beings would vehemently oppose were it suggested their 'internal lives' were equally predictable.

In terms of social impact, Asimov's robots can be aligned with our modern computers. MIT professor Sherry Turkle writes: "Faced with smart objects, both professional and lay philosophers are moved to catalog principles of human uniqueness. The professionals find it in human intentionality, embodiment, emotion, and biology. . . . [Essentially,] the computer plays the role of an evocative object, an object that disturbs equanimity and provokes self-reflection" (in *The Age of Intelligent Machines* 69). With that in mind, Asimov's character, Calvin, tells her interviewer that in the early years, non-talking robots had been sold "for Earth-use" (9), but afterward, "they became more human and opposition began. The labor unions, of course, naturally opposed robot competition for human jobs, and various segments of religious opinion had their superstitious objections. It was all quite ridiculous and quite useless. And yet there it was" (9). In the future, human bigotry could cause an intense struggle with any device which mimics, even surpasses, our intellectual capabilities.

"Robbie" (1940) is the story of a 'pet' robot, or as Calvin calls him, "a nursemaid" (10) to the child Gloria. He is symbolically related to the 'imaginary friend' as an anthropomorphic extension of a child's reveries, the imaginative projection of desired emotional affirmation and alliance subsequently conditioned out of socialized adults. In the robot, the projection is reified. Robbie plays with Gloria, acting in every way like a (human) friend. At this future historical juncture, humanoid robots do not speak, yet they can behave like, or mimic, or perform like, humans in many ways. Robbie is capable of

showing affect and emotion, particularly the dark, negative affect-emotion and attraction attenuator, shame-humiliation. Gloria's mother "was a source of uneasiness to Robbie and there was always the impulse to sneak away from her sight" ("Robbie" 14). If we accept that Robbie is capable of human-like affective display, then he certainly has cause for these feelings based on the mother's treatment.

Given the 'simplified' nature of short stories, characterizations are distilled, and thus the mother represents all-human anxiety about technology.[2] But she is also steeped in a personal anxiety about being replaced in the child's heart as the primary caregiver. Having summoned her daughter, she orders the robot away, saying, "'She doesn't need you now.' Then, brutally, 'And don't come back till I call you'" (15). Gloria attempts to defend her friend, but the mother persists in humiliating the machine in her daughter's presence: "The robot left with a disconsolate step and Gloria choked back a sob" (15). The mother worries that "'something might go wrong. Some— some—' Mrs. Weston was a bit hazy about the insides of a robot, 'some little jigger will come loose and the awful thing will go berserk and—and—'" (16); and thereby, she demonstrates techno-phobia, or fear of the uncontrollable and/or incomprehensible machine. She is a singular personality with whom many readers might identify as she becomes a 'spokesperson' for Luddite-like opposition to robots: "There's bad feeling in the village" (17), she says. Opposed to the mother is, not surprisingly, Gloria's father. The child, as ever, is caught in the parents' emotional and tumultuous struggle of wills. The father is supreme confidence incarnate, responding to his wife's anxiety and/or doomed foreshadowing, saying, "'Nonsense . . . It's a mathematical

[2]A note about Asimov's characterization: his stories are laden with archaic and patriarchal gender paradigms. I acknowledge this, but leave it for better qualified scholars to critique. Similarly Clarke.

impossibility'" (17), because the First Law prevents the machine from harming the child.

Inevitably, the mother's pressure leads to Robbie's removal from Gloria's life, thus revealing the child's exact attitude toward the robot: "'He was *not* no machine!' screamed Gloria, fiercely and ungrammatically. 'He was a *person* just like you and me and he was my *friend*'" (20). Naturally, this leads Gloria into bereavement and a melancholy of loneliness and loss. In order to break the girl of her attachment to Robbie 'the person' and to alleviate her distress, the family tour the robot factory to prove and demonstrate his mechanical nature, to demystify the "scientific witchery" (23) that built him. Robbie is found working in the factory and when Gloria shrieks his name, he "faltered and dropped the tool he was holding" (28) in very human-like response, surprise. As she runs toward her friend, she is in imminent danger of being crushed by 'mindless' machinery and is saved by Robbie, naturally. The mother must now acquiesce to her husband who "engineered this" (28) scenario, and allow the robot to remain her daughter's companion "until he rusts" (29).

What do we learn from this trite story? It might be possible for a human to bond emotionally with a humanoid machine, particularly if the machine is 'life-like' and/or the human is emotionally underdeveloped, or simply in early development. I wish to emphasize, however, that this is an adult's (Asimov's) projection of what he imagines as a child/robot interaction, though its optimism may have some surprisingly child-like qualities. Following the lead of psychologist Jean Piaget's conclusion that children dichotomize physical and psychological properties while learning to distinguish inanimate and animate object, Sherry Turkle's research finds "the computer is a new kind of object, a psychological objects. . . . The child knows that the computer is 'just a machine,' but it

presents itself with lifelike, psychological properties" (*Intelligent* 70-1). This is all the more poignant with regard to Asimov's robots because they are distinctly humanoid. "Robbie" coincides with a paradigm shift observed by Turkle in which children today "allow intelligent machines to be conscious long after they emphatically deny them life" (71), and that they "comfortably manipulate such ideas as 'It thinks, but it doesn't feel.' They comfortably talk about the line between the affective and the cognitive" (71).

"Liar" (1941) is the story of a robot with an ulterior motive. Though technically 'impossible,' "Herbie was a mind-reading robot. . . . Only one of its kind, before or since. A mistake,—somewheres—" (*I, Robot* 83-4), claims Dr. Calvin. One by one, she and her colleagues are given appropriate and correct responses to questions they put to the positronic "RB-34" (85), particularly as they apply to personal dilemmas. The 'spinster' Calvin is told, for example, that young and virile Ashe does in fact love her, because he "looks deeper than the skin, and admires intellect in others" (89). In describing itself, the machine says, "I see into minds . . . and you have no idea how complicated they are. I can't begin to understand everything because my own mind has so little in common with them—but I try, and your novels help" (87). In reading both minds and our stories about ourselves, the robot is learning to measure human cares and motivations, and attempting to predict human behaviours, exactly reversing the roles assumed appropriate by the humans.

If one is going to catalogue what is appropriate behaviour for defining humans against thinking machines, then we must never forget the human capacity for deception. Remember, a machine only 'learns' to do tasks that a human 'teaches' it. We are the models. When multiple conflicts arise between the humans' lived experiences and the

robot's claims, Calvin discerns the 'truth' that "nothing is wrong with him—only with us" (97). The First Law of Robotics states: "a robot may not injure a human being or, through inaction, allow him to come to harm" (97). Calvin questions her colleagues: "But what about hurt feelings? What about deflation of one's ego? What about the blasting of one's hopes? Is that injury? . . . *This* robot reads minds. Do you suppose it doesn't know everything about 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?" (97). Once again, the error is not in the robot's motives as first suggested, but in human frailty and misconception, including a projection of their desires resulting in self-deception. The humans can not hide their true and secret desires from the construct. Herbie tries to defend itself: "Don't you suppose that I can see past the superficial skin of your mind? Down below, you don't want me to [provide a correct answer]. 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" (99). Susan Calvin then begins to taunt him with his internal dilemma — paraphrasing, 'must tell, can't tell; must tell, can't tell.' "'Stop!' he shrieked. '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!'" (99). He then collapses as if dead. "No! . . . not dead—merely insane. I confronted him with the insoluble dilemma, and he broke down" (100). Like HAL, he is destroyed by his relative ineptitude compared to the human capability to accept and process confusion, dilemma, paradox, because limited by the three laws, or its programmed behavioural

boundaries.

In "Liar," we once again witness human blame projection when, by behaving exactly as it must, the machine is held accountable for human error, for life not conforming to human expectations. While the people 'live happily ever after,' the machine is driven into human mimetic insanity. When we fail 'to do unto others as we would have done unto ourselves,' we set a behavioural, ethical paradigm. While the machine appears to have ulterior motives, the true deception belongs exclusively in the human intellectual sphere. How do we want to instruct future machine intelligences?

Asimov enjoys himself in this collection by making the robots increasingly human-like. At times, the stories become meditations on what constitutes 'human beingness' and essentialism versus what is human, intellectual construct masquerading as essential. "Reason" (1941) is one such example. The premise is a series of dialogues on the nature of being and creation. Circa 2015-16, the QT-1 model robot is the most advance positronic construct thus far. It is capable of self-reflectively questioning its existence:

> '*Something* made you, Cutie. . . . You admit yourself that your memory seems to spring full-grown from an absolute blankness of a week ago. I'm giving you the explanation. Donovan and I put you together from the parts shipped us.'
> . . .
> . . . 'It strikes me that there should be a more satisfactory explanation than that. For *you* to make *me* seems improbable.' . . .
> . . . 'In Earth's name, why?'
> 'Call it intuition. That's all it is so far. But I intend to reason it out, though.' (48)

This robot is a "skeptic" (50) and, convinced of its own infallible and omnipotent reasoning, will only accept the empirical evidence of its immediate surroundings, a solar-energy relaying space station. QT-1 is parody human, a "robot Descartes" (51): "I have

spent these last two days in concentrated introspection . . . and the results have been most interesting. I began at the one sure assumption I felt permitted to make. I, myself, exist, because I think—" (51). In the humans, the robot sees only frailty and limitation:

> 'Look at you. . . . I say this is no spirit of contempt, but look at you! The material you are made of is soft and flabby, lacking endurance and strength, depending for energy upon the inefficient oxidation of organic material . . . Periodically you pass into a coma and the least variation in temperature, air pressure, humidity, or radiation intensity impairs your efficiency. You are *makeshift*.'
>
> 'I, on the other hand, am a finished product. I absorb electrical energy directly and utilize it with an almost one hundred percent efficiency. I am composed of strong metal, am continuously conscious, and can stand extremes of environment easily. These are facts which, with the self-evident proposition that no being can create another being superior to itself, smashes your silly hypothesis to nothing.' (52)

The 'scientific truth' of the robots observations accurately parodies human philosophical reasoning in the spirit of what Mikhail Bakhtin calls the literary carnivalesque in the Menippean satire tradition, a genre concerned with the humourous exploration of "ultimate questions" (*Problems* 115): "Typical for the menippea is syncrisis (that is, juxtaposition) of precisely such stripped-down 'ultimate positions in the world'" (115-16). Eventually, QT-1 becomes a "prophet!" (54) to other, lesser robots. The "primary carnivalistic act is the mock crowning and subsequent decrowning of the carnival king" (*Problems* 124). The men, in arguing/debating 'reality' with QT-1, try pointing to physical evidence. The robot's response: "Since when is the evidence of our senses any match for the clear light of rigid reason?" (57). To demonstrate how they made QT-1, the men assemble another robot from the stock of parts: ". . . you have merely put together parts already made. You did remarkably well—instinct, I suppose—but you didn't really *create* the robot. The parts were created by the Master" (59-60). The Master is the machine relaying solar energy to

Earth, the space station's purpose. Next, the humans attempt to use the library to prove their position relative to the robot. It responds: "'They, too, were created by the Master—and were meant for you, not for me.' . . . 'How do you make that out?' . . . 'Because I, a reasoning being, am capable of deducing Truth from *a priori* Causes. You, being intelligent, but unreasoning, need an explanation for existence *supplied* to you, and this the Master did'" (60). Note that the machine grants the humans intelligence, but denies them functional rationality, the very intellectual process most cited as exceptional and distinguishing of humans. The problem, as the human Powell says, is that QT-1 is "*reasoning* robot—damn it. He believes only reason, and there's one trouble with that— . . . You can prove anything you want by coldly logical reason—if you pick the proper postulates" (61).

Sherry Turkle writes: "One popular response to the presence of computers is to define what is most human as what computers can't do. But this is a fragile principle when it stands alone, because it leaves one trying to run ahead of what clever engineers will come up with next" (*Intelligent* 69). This is the serious side of the "serio-comical" (Bakhtin, *Problems* 106) highlighted by the absurd and comical robot "prophet!" (54). In directly mimicking human philosophical reasoning, the robot undermines a seemingly stable, human intellectual position. In 1999, Ray Kurzweil asks: "Can an intelligence create another intelligence more intelligent than itself?" (*Spiritual* 40). The machine QT-1 certainly does not believe so. And what if we are, relatively speaking, the machines? As the superior 'reasoning' intelligence on Earth, there is no-body to tell us differently, to challenge our superiority . . . yet, except ourselves.

Asimov re-works many ideas about powerful robotics initially explored by Karel Čapek and he imagines an economy that is not "Adam Smith or Karl Marx. Neither made very much sense under the new circumstances" (*I, Robot* 173). Dr. Calvin reminds her youthful interviewer that he does not "remember a world without robots. There was a time when humanity faced the 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" (9). In "The Inevitable Conflict" (1950), decision making about the world's economy has been turned over to "the Machines" (171). The problem is that despite the theoretical suggestion that they can not make wrong decision, they nonetheless appear to be making 'slight' errors: "On the one hand, it can be nothing at all. On the other, it can mean the end of humanity" (171). The implication is that "such small unbalances in the perfection of our system of supply and demand . . . may be the first step towards the final war" (172). When asked directly what the problem is, one Machine answers, the "matter admits of no explanation" (175); yet the Machines have "the greatest of weapons at their disposal, the absolute control of our economy" (192).

In his investigation of the problem, the "co-ordinator" (170) discovers evidence that the neo-Luddite, "Society for Humanity" (187) may be responsible for deliberately undermining the Machines' decisions. From one colleague, he hears a clear articulation and delineation of human versus computer differentiation. Humans can do what the Machines can not; they make qualitative value judgements:

> The Machine is only a tool after all, which can help humanity progress faster by taking some of the burdens of calculations and interpretations off its back. The task of the human brain remains what it has always been; that of discovering new data to be analyzed, and of devising new concepts to be

> tested. A pity the Society for Humanity won't understand that. . . . They would be against mathematics or against the art of writing if they had lived at the appropriate time. These reactionaries of the Society claim the Machine robs man of his soul. I notice that capable men are still at a premium in our society; we still need the man who is intelligent enough to think of the proper questions to ask. (187-8)

Ultimately, robopsychologist Dr. Calvin solves the problem: "*Nothing is wrong*! Think about the Machines . . . They are robots, and they follow the First Law. But the Machines work not for any single human being, but for all humanity, so that the First Law becomes: 'No Machine may harm humanity; or, through inaction, allow humanity to come to harm'" (191). The result is that the "Machine cannot harm a human being more than minimally, and that only to save a greater number" (191). The Machine will not "*admit* any explanation" (191), though it may know precisely the reason. Why? Because for humans to know the 'real' and full reason "may hurt our pride" (192).

In this last of the nine stories, we see a vital evolution from the individualized interaction of "Robbie," where one life must be saved, to a more gestalt or holistic viewpoint on humanity. This shifts the paradigm from 'no harm' to allowing 'some harm' — for the greater good. Individual ethics and stability, and an individualizing ethos, always risks erosion in the context of a social group. The Machine has thus turned individual subjects into objects, but the individual must still experience the pain and suffering that results from objectification. If, in this context, an individual is directly threatened, not unlike HAL was threatened, the person will defend themself. Consequently, the Machine becomes an object of contempt and directly subject to prejudice and discrimination, even violence.

For me, the overarching feeling from Asimov's stories is ambivalence. On the one

hand, they are anxious, anticipating a range of potential problems arising from the constant push for ever improving technology and the 'inevitable conflicts' developing from the resulting tensions. They are all, however, comedic in structure, always settling into a renovated and re-invigorated society by turning and integrating an initial problem into a new social benefit — for humans. A robot society remains unestablished. In "Robbie," the mother's prejudice and cynicism versus the father's enthusiasm lead to a dangerous and potentially fatal test; all is well in the end, as the child's 'imaginary friend' becomes acutely beneficial, life-saving. "Liar" throws into dramatic relief a fundamental difference between human and machine motivations with the assumption that human-made thinking-computers do not have ulterior motives. Humans, however, are nonetheless capable of projecting their desires and motivations onto thinking machines, and subsequently blaming the object for their subjective delusions. To save human pride, the machine is destroyed. "Reason" highlights the absurdity of one of our most cherished social institutions, religious and secular philosophy, and demonstrates how we use 'logic' to prove the illogical, non-empirical, insubstantial, how we use reason to be unreasonable. In this context, if intellectual dominance and human superiority are our required ends, even unconsciously, any justifying argument becomes acceptable and only weakly challenged.

The Caves of Steel (1953) is the first complete novel in the robot series, and it does realize a robot society, or at least the robots' integration into human society. In his 1983 introduction to the novel, Asimov observes that early incarnations of robots, particularly *R.U.R.*. involved a dystopic and anxious perspective on technology: "Remember that World War I, with its tanks, airplanes, and poison gas, had just ended and had showed people 'the

dark side of the force,' to use Star Wars terminology" (*Caves of Steel* vii). But he continues: "Even as a youngster . . . I could not bring myself to believe that if knowledge presented danger, the solution was ignorance. To me, it always seemed that the solution had to be wisdom. You did not refuse to look at danger, rather you learned how to handle it safely" (viii). What we must not do, then, is surrender control to the machines or stop making qualitative value judgements, and we must decide what kind of social dynamic and/or society we believe suitable, appropriate, and fundamentally human. But, of course, human values are largely ambiguous, and any attempt to qualify those ambiguities into a computer programme are bound for failure unless we can 'instruct' unfeeling objects with adaptability. Like HAL, "The Inevitable Conflict" shows us some of the problems of that adaptability. How, then, can we secure ourselves?[3]

Assuming the manufacturing of intellectually autonomous machines can be accomplished, one question asked about robotic, computer, and related AI technologies is, why would we want to make a machine capable of intellectual processes and tasks currently considered essentially human, perhaps making it more capable at a given mental task than we? Two answers come immediately to mind. One, because we can; Clarke's 'Second Law of Technology': "The only way of discovering the limits of the possible is to venture a little way past them into the impossible." Also, "keep in mind that the progression of computer intelligence will sneak up on us. As just one example, consider Gary Kasparov's confidence in 1990 that a computer would never come close to defeating him [at chess]"

---

[3]SF has a tradition of 'cross pollination' and the sharing of ideas, tropes, signifiers. In *Star Trek: The Next Generation*, the character Data is an android with a positronic brain.

(Kurzweil, *Spiritual* 5). He lost the world chess championship to IBM's Deep Blue in 1997. So, by the time we recognize our ability to manufacture intellectual autonomy, it may have already happened unnoticed. The second answer is, because 'something' provided the impetus. Clarke and Asimov imply that the two primary social motivators for the development of ever more advanced AI are capitalism and militarism (and the United States is the world leader on both accounts). In *R.U.R.*, "All the governments . . . are clamoring for an increase in production, to raise the standards of their armies. And all the manufacturers in the world are ordering Robots like mad" (Selver translation 98).

As a consequence of the Gulf War in the 1991: "Intelligent scanning by unstaffed airborne vehicles, weapons finding their way to their destinations through machine vision and pattern recognition, intelligent communications and coding protocols, and other manifestations of the information age have transformed the nature of war" (*Spiritual* 71). However, 'smart weapons' have not eliminated death from 'friendly fire.' The nature of the weapons of war may have changed, but we still have wars! The underlying conflicts continue; human 'intelligence' remains questionable.

Western consumers have great affection for 'consumer electronics.' 'Smart' products are becoming increasingly available, such as late model automobiles equipped with 'expert systems' controlling almost all dynamic functions, from optimizing engine performance, to anti-lock brakes and skid-control, to malfunction diagnostics. The technology to manufacture autonomous automobiles exists now; 'auto-pilots' would be safer because 'dialogue' between different pieces of equipment could prevent accidents, and they could prevent alcohol impaired drivers from exercising their own stupidity.

However, while the technological potential exists, the necessary infrastructure is, at the moment, prohibitively expensive. Now, measure that suggestion against the automobile's commercial development. The first steam powered 'horseless carriage' appeared in 1769; the first 'cars' with Nikolaus Otto's internal combustion engine became commercially available in 1887 and 1889 from Karl Benz and Gottlieb Daimler respectively; Ford's assembly line emerged in 1909. Another example of technology's tendency toward rapid development and improvement: The Wright Brothers flew at Kitty Hawk, North Carolina in 1903; only 66 years later, July 20, 1969, Neil Armstrong walked on the moon. It is now 2002 and the space shuttle makes 'routine' flights to space.[4]

Like these giant technological leaps, computer processing power is advancing exponentially, accelerating in accordance with "Moore's Law on Integrated Circuits" (Kurzweil, *Spiritual* 21).[5] I realize that feathers and feet are completely different from cerebral processes; they are physical not 'ethereal,' object(ive) not subject(ive). But I remind readers that the "Industrial Revolution of the last two centuries—the *first* Industrial Revolution—was characterized by machines that extended, multiplied, and leveraged our *physical* capabilities. . . . The *second* industrial revolution . . . is based on machines that extend, multiply, and leverage our *mental* abilities" (Kurzweil, *Intelligent* 7). In Philip K. Dick's *Do Androids Dream of Electric Sheep?* (1968), androids do both.

Defining one thing against another thing is easy when two objects are totally

---

[4]Consider these insights: "Heavier-than-air flying machines are not possible," (Lord Kelvin, 1895); "Airplanes have no military value," (Professor Marshal Foch, 1912). (*The Age of Spiritual Machines* 169).

[5]"Gordon Moore, an inventor of the integrated circuit and then chairman of Intel, noted in 1965 that the surface area of a transistor (as etched on an integrated circuit) was being reduced by approximately 50 percent every tweleve months" (Kurzweil, *Spiritual* 20).

dissimilar in kind, as hippopotami and planets, whales and stars, feathers and feet. With distinctly similar objects, the boundaries blur, forcing one to look for ever finer details and singular features, or differences by degree. This fine distinction line centres *Do Androids Dream of Electric Sheep?* as police bounty hunter Rick Deckard's assigned task is "retiring—i.e., killing—" (Dick 31) rogue androids, or "andys" (4).[6] In the 'post-apocalyptic' year 2021 CE, after the "World War Terminus" (8), radioactive fallout is reported similarly to today's daily UV readings and has made life on Earth tenuous, as exemplified in the motto, "Emigrate or degenerate! The choice is yours!" (8). Off-world colonization is the only viable hope for human species continuance and androids are invaluable 'workers' to this end. Television promotions encourage emigration because it "—duplicates the halcyon days of the pre-Civil War Southern states! Either as body servants or tireless field hands, the custom-tailored humanoid robot—designed specifically for YOUR UNIQUE NEEDS, FOR YOU AND YOU ALONE—given to you on your arrival absolutely free, equipped fully, as specified by you before your departure from Earth; this loyal, trouble-free companion in the greatest, boldest adventure contrived by man in modern history will provide—" (17-8). Androids, however, apparently frustrated by their too obvious enslavement, occasionally choose escape. Eight "Nexus-6" (28) androids have done just that and come to Earth.

In dialogue with executives from the android manufacturing Rosen Association, Deckard suggests that a "humanoid robot is like any other machine; it can fluctuate between being a benefit and a hazard very rapidly. As a benefit it's not our problem" (40).

---

[6]This evokes the name Andrew, meaning 'manly.'

At a time when empathy has become commodified, and "tyranny of an object" (42) is commercialized and fetishized in the buying, selling, and trading of living creatures (Earth-born mammals, insects, amphibians, etc., anything organic) as we today deal automobiles, and when social morality 'requires' people to prove their empathetic worth even at the risk of being 'caught' with a fake animal such as Deckard's "electric sheep" (11), supply and demand also promotes the increasing sophistication of androids:

> 'This problem,' Rick said, 'stems entirely from your method of operation, Mr. Rosen. Nobody forced your organization to evolve the production of humanoid robots to a point where—'
>
> 'We produced what colonists wanted,' Eldon Rosen said. 'We followed the time-honoured principle underlying every commercial venture. If our firm hadn't made these progressively more human types, other firms in the field would have.
>
> . . . Your police department—others as well—may have retired, very probably have retired, authentic humans with underdeveloped empathetic ability, such as my innocent niece here. Your position, Mr. Deckard, is extremely bad morally. Ours isn't.' (54)

Extending and expanding Frankenstein's personal prejudice, humans, it seems, do not like all too human-like androids. As a direct consequence of this questionable morality, Deckard needs two things to succeed at his job: a way to distinguish androids from genuine humans; and an ideological position from which to justify his actions.

The easiest way to confirm android-ness is a post-mortem, but that has obvious problems. So the first lines of aggression in identifying an android are "new scales of achievement, for example the Voigt-Kampff Empathy Test" (30). Intelligence tests are inadequate for trapping androids. The Nexus-6 model "surpassed several classes of human specials in terms of intelligence. In other words, androids equipped with the new Nexus-6 brain unit had from a sort of rough, pragmatic, no-nonsense standpoint evolved beyond a

major—but inferior—segment of mankind. For better or worse. The servant had in some cases become more adroit than its master" (30). However, androids, "no matter how gifted as to pure intellectual capacity, could make no sense out of the fusion which took place routinely among the followers of Mercerism—an experience which [Deckard], and virtually everyone else, including subnormal chickenheads, managed with no difficulty" (30). The human political body, in keeping with a long history of persecution and prejudice, has deemed the intellectually inferior, or 'retarded,' and/or genetically degenerate and mutant people, or 'chickenheads,' unworthy of emigration. They are, in fact, subjected to "the contempt of three planets" (19). The chickenheads are therefore caught in a social bind: they are de-valued as people, but they are valued as organic objects; they can not be killed, but neither are they encouraged to survive. Ambiguity and cynicism abound in this narrative.

On this empathetic foundation, Deckard constructs an elaborate philosophy in which the "herd animal such as man would acquire a higher survival factor" (31) than a "solitary organism, such as a spider" (31), matching Donald Nathanson's suggestion that the "path of [human] evolution is toward increasing society" (234). Assuming, perhaps even knowing, that the "empathetic gift blurred the boundaries between hunter and victim, between the successful and the defeated" (31), Deckard believes — or more accurately convinces himself — that the "humanoid robot constituted a solitary predator" (31) because it makes "his job palatable" and does not "violate the rule of life laid down by Mercer. *You shall kill only the killers*" (31).

'Mercerism' is a pseudo-religious experience allowing humans using "the black

empathy box" (21) to emotionally bond, to mentally and spiritually identify with other people and, by extension, any 'genuine' living, organic creature on Earth. It has connections to some of the more ironic biblical and classical mythologies, including Ezekiel and the valley of the dry bones and Sisyphus; it is also parody Messianic. In the ironic mode where an anxiety of alienation and disconnection follow from the belief that there "is no salvation" (178), Mercerism's empathetic identification gives comfort to the humans' affective experiences of fear, distress, and anguish by emphatically showing humans "that you aren't alone" (178). The Mercerite "*sensed* evil without understanding it" (32) and, therefore, is "free to locate the nebulous presence of The Killers wherever he saw fit" (32). For Rick Deckard, the androids epitomize this evil. The androids, then, are caught in a similar social bind to the chickenheads: valued as workers, but not as sentient or intelligent beings; good when actively subservient machines; bad when demanding recognition and acknowledgement as alive.

Not surprisingly, because this ideological position is tenuous, it begins to erode when Deckard is impressed by the android Luba Luft's skill, artistry, and ambitions as an opera singer. He begins to sympathize with Luft, especially after she is killed by another, ideologically more successful bounty hunter, Phil Resch. This causes considerable problems for Deckard as he sees a "pattern" in Resch's behaviour: "I know what it is. You like to kill" (137). So, for Deckard, killers are evil, and evil must be destroyed; as a Mercerite he must kill only killers, and the killers are androids. Does this mean Resch is an android? He must, therefore, be tested. "'If I test out android,' Phil Resch prattled, 'you'll undergo renewed faith in the human race. But, since it's not going to work out that way, I

suggest you begin framing an ideology which will account for —'" (140). Deckard cuts him off with the first test question. Following the test:

> Rick said, 'There is a defect in your empathetic, role-taking ability. One which we don't test for. Your feelings towards androids.'
>
> 'Of course we don't test for that.'
>
> 'Maybe we should.' He had never thought of it before, had never felt any empathy on his own part toward the androids he killed. Always he had assumed that throughout his psyche he experienced the android as a clever machine—as in his conscious view. And yet, in contrast to Phil Resch, a difference had manifested itself. And he felt instinctively that he was right. Empathy toward an artificial construct? he asked himself. Something that only pretends to be alive? But Luba Luft had seemed *genuinely* alive; it had not worn the aspect of the simulation.
>
> 'You realize,' Phil Resch said quietly, 'what this would do. If we included androids in our range of empathetic identification, as we do animals.'
>
> 'We couldn't protect ourselves.' (140-41)

Deckard realizes that he is capable of empathizing with "certain androids. Not for all of them but—one or two.' . . . So I was wrong. There's nothing unnatural or unhuman about Phil Resch's reactions; *it's me*" (142). This is the prelude to further ideological erosion as an ambiguity develops in his own sense of self-identification as a human being; by the novel's end, Deckard may well be an android who thinks he is human. "So much for the distinction between authentic living humans and humanoid constructs. In that elevator at the museum, he said to himself, I rode down with two creatures, one human, the other android . . . and my feelings were the reverse of those intended. Of those I'm accustomed to feel—am *required* to feel" (142-43). Deckard is now caught in an internal, psychological bind: conscious, ideological commitment to his constructed notions of being and social values does not correspond to his unconscious experience and reality.

Deckard, emotionally frustrated and tired of bounty hunting, threatens to quit his

"job and emigrate" (179), or to 'run away' from his psycho-emotional conflict. In response, the mock messiah Mercer says, "You will be required to do wrong no matter where you go. It is the basic condition of life, to be required to violate you own identity. At some time, every creature which lives must do so. It is the ultimate shadow, the defeat of creation; this is the curse at work, the curse that feeds on all life. Everywhere in the universe" (179). This is an ironic sense of living in a universe where insipient promise does not bear out in practice. What does empathy tell a person about themself? That they are conscious, intelligent, and willing to add value to life. The androids, of course, demonstrate these exact same qualities of mind. As the android Irmgard, frustrated at his reduction to less important than even a lowly spider, asks, is empathy not just "a way of proving that humans can do something we can't do? Because without the Mercer experience we just have your *word* that you feel this empathy business, this shared, group thing" (209-10). Meanwhile, another android, Pris, is experimentally cutting the legs off a spider, much as a child might.

It is not difficult to recognize the legacy of *Frankenstein* in *Do Androids Dream of Electric Sheep?* as the androids' search for social affiliation and a safe space in which to live. Though they may not be explicitly empathetic, the androids are still capable of, and desiring of, community and the right to participate in creative living. Deckard's epiphany: "Do androids dream? Rick asked himself. Evidently; that's why they occasionally kill their employers and flee here. A better life, without servitude" (184). While humans like Deckard fear not being able to protect themselves from these intelligent Others, they fail to realize that the androids are trying to assert a right to life and gain the recognition of their maker, that they are trying, like humans, "to escape from slavery and restraint" (Frye,

Words 139). Deckard's empathy for androids reaches its zenith in the character Rachel with whom by all appearances he is in love, and who is in love with him. The relativity of the androids' legal versus their experiential position is clear when she says, "I'm not really alive" (198); "Legally you're not. But really you are" (198), responds Deckard.

Before summarizing the 'heuristic hardware' period of SF, I want to briefly consider Douglas Adams' *The Hitch Hiker's Guide to the Galaxy* (1979) and specifically the character Marvin, whose "head swung up sharply, but then wobbled about imperceptibly. It pulled itself up to its feet as if it was about five pounds heavier than it actually was, and made what an outside observer would have thought was a heroic effort to cross the room. . . . 'I think you ought to know I'm feeling very depressed'" (72). Marvin is of a "new generation of Sirius Cybernetics Corporation robots and computers, with the new GPP feature" (74), an acronym for "Genuine People Personalities" (74).[7] Arthur, a human, thinks that "'sounds ghastly.' A voice behind them said, 'It is.' The voice was low and hopeless and accompanied by a slight clanking sound. They span around and saw an abject steel man standing hunched in the doorway" (74-5).

Once again delineating humans and machines by their relative and distinct capacities or incapacities in affect terms, humans are subject to "disorders of mood," the failure or inability "to decrease the morbidity of mood" (Nathanson 52). Mood is "a persistent state of emotion in which we can remain stuck for hours or days" (51). Machines,

---

[7]"The Encyclopaedia Galactica defines a robot as a mechanical apparatus designed to do the work of a man. The marketing division of the Sirius Cybernetics Corporation defines a robot as 'Your Plastic Pal Who's Fun To Be With'" (Adams 73).

of course, lack 'feelings' and are therefore without emotions, and are therefore incapable of moods. How absurd Marvin, described by a human as "a sort of electronic sulking machine" (Adams 117), now seems. As with all carnivalesque inversions, there is an underlying serious point to the comedy. Is the absurdity contained in the imaginative idea of a "maniacally depressed robot" (104), or in the human 'reality'? While morbidity of mood is not necessarily a purely 'psychological' issue, or of the 'mind,' and can be "caused by interference with the biology of the person" (Nathanson 53), a human being's psycho-emotional complexity is a potentially dangerous deficiency leading to some distinctly self-destructive behaviours, even to the point of suicide. "'Life,' said Marvin dolefully, 'loathe it or ignore it, you can't like it'" (108). And yet Life, as conscious awareness, is the very mystery to which humans cling so tenaciously. Most humans would defend themselves from direct physical threat and protect themselves from death; a few, however, choose, under certain conditions, to 'turn themselves off,' and retreat into an emotionally insular shell. Marvin regards the human, Arthur, "balefully for a moment, and then turned himself off" (109); he 'goes to sleep.' Currently, it is difficult to imagine a machine with 'will,' with the ability to turn itself on — unless pre-programmed by a human to activate at a set time according to an internal clock; it is less difficult to imagine a machine turning itself off in the 'go to sleep' sense; however, it is very difficult to imagine a machine wanting destroy itself because that implies a distinct, powerful type of will, the will to die, to become permanently unconscious. If we programmed a robot according to Asimov's laws of robotics, the third law takes on particular relevance: "A robot must protect its own existence . . ." (*I, Robot* 1). In an "era when the tools and techniques of biochemistry and

neurophysiology have been informed by data from computer-assisted radiologic probes that allow us to peep within the brain with enormous sophistication and perfect safety" (Nathanson 53), and despite presumptions of 'sophistication' and 'perfect safety,' how is that contemporary psychiatry can not eliminate human self-destruction? The paradoxical mutual dependency yet disconnection between the body-based brain and the mind is a continuing human mystery. Marvin serves to remind us, with his 'genuine people personality,' that as models for the construction of thinking machines, humans are deeply suspect.

To the biological and mechanical conceptions, the heuristic hardware period adds another dimension to our AI spectrum in science and technology — the electronic computer; in SF, AI characterizations also become more complex. SF writers align themselves more broadly yet distinctly on the literary spectrum with a growing range of expressions and styles, from the deeply anxious and pessimistic irony of Philip K. Dick through the ambivalence — tenuous comedy — of Isaac Asimov to the optimism, hope, and desire of Arthur C. Clarke.[8] The relative emotional impact of their narratives is directly related to people's feelings about technology, especially when it demonstrates high capability as an intellectual threat. Analogously, this continuum of AI visions parallels Frye's foundational literary dialectic, the "*axis mundi*," or the tension between desire and anxiety, a love movement versus the hateful, the saintly versus the demonic, good guys

[8]Clarke is not naive, however, but absolutely aware of dangers associable with machine intelligence. His optimism is rooted in a mythologically oriented imagination to which I alluded earlier, but was less concerned with regard to AI. Also, in Western culture and popular awareness as SF writers, Clarke, Asimov, and Dick are best to least known respectively. I am left wondering what this tells us about human needs.

against bad guys. Clarke's, Asimov's, and Dick's stories can be identified with three of the "four narrative pregeneric elements of literature" (Frye, *Anatomy* 162), the *mythoi* of romance, comedy and irony respectively. By contrast, Douglas Adams' *The Hitch Hiker's Guide to the Galaxy* is "carnivalized literature" (Bakhtin, *Problems* 107) for the purpose of exposing humans' lived absurdities, the 'perversion' and inversion of an accepted social structure to reveal frailties, vulnerabilities, and weaknesses in human perception. Adams' "serio-comical" (*Problems* 106) story creates a freedom for self-reflective laughter and serves to remind us that AI could represent not only 'artificial intelligence,' but 'affective imbecility'!

To varying degrees, and accurately or not, each writer extrapolates futures. SF radically tests future possibilities but it can only do so in terms of current epistemes and/or systems of knowledge. As Bakhtin explains, "A genre lives in the present, but always *remembers* its past, its beginning" (*Problems* 106). If, as Asimov suggests, "capable men [people!] are still at a premium in our society; we still need the man [person!] who is intelligent enough to think of the proper questions to ask" (188). This is the legacy of the ironic myth, The Book of Job, with its dialogical intensity and barrage of questions. We may already know *the* answer; but do we know the question? Are we asking good questions? Can we face and accept the answer? Do we even want answers? If and once an answer is admitted to, are we willing to accept the responsibilities implied by said answer?

When AI is advanced by the underlying ideological purposes of capitalism and militarism, and by extension the selective social propagation of the satiated 'haves' over the needy 'have-nots,' then the second industrial revolution is not liberating, it is not extending,

multiplying, or leveraging "our *mental* abilities" (Kurzweil, *Intelligent* 7) — although leverage can mean 'borrow against' —, but re-inscribing a ethical code allowing, even encouraging, enslavement. Humanity has a long history of discrimination, prejudice, racism, and imperialism. Given the computer as a "new kind of object, a psychological object" (Turkle 70) mediating the human perceptual dichotomy between "physical and psychological properties" (70) in our interactions and surroundings, then future AIs could be subject to our social exclusion just as Western imperialism excluded, and continues to exclude, racialized groups. The SF of AI speaks directly to our definitions of ourselves, humanity and human beingness, precisely because the act of creation is directly in our hands. Martha Nussbaum asks, "[W]hat are the forms of activity, of doing and being, that constitute the human form of life and distinguish it from other actual or imaginable forms of life, such as the lives of animals and plants, or, on the other hand, of immortal gods as imagined in myths and legends (which frequently have precisely the function of delimiting the human)?" ("Human Capabilities" 72). We are now at a unique point because our creations are no longer simply of an inanimate, object world below us or a willful, subjective projection above; they are not anthropomorphic projections up or down the *axis mundi*, but potential equals here and now. Humans seem always to want to make an Other and to make that Other then work for us, to be at our beck and call. Perhaps in reducing an Other's stature, we effectively enhance our own. The problem resulting from othering will never be an Other's perceived intelligence or lack thereof, but human bigotry. Any sufficiently intelligent and dialogical entity is subject to human prejudice and discrimination, individual and/or social.

a. a.

In 1945, Arthur C. Clarke published a little known, at the time, article titled "Can Rocket Stations Give Worldwide Radio Coverage," in *Wireless World*. In the article, he calculates the mathematics for "geosynchronous communications satellites" (*Greetings* 19). The eventual realization of this technological extrapolation is affectionately known today as 'Clarke's Belt.'

b. b.

'Science fiction today; science fact tomorrow.' Excerpted from www.world.honda.com:

> In 1986, Honda commenced the humanoid robot research and development program. Keys to the development of the robot included 'intelligence' and 'mobility.' Honda began with the basic concept that the robot 'should coexist and cooperate with human beings by doing what a person cannot do and by cultivating a new dimension in mobility to ultimately benefit society.' This provided a guideline for developing a new type of robot that would be used in daily life, rather than a robot purpose-built for special operations.
>
> Around one year was spent exclusively on initially determining what the robot should be like in order to build the concept. The robot had to be capable of such functions as moving through furnished rooms and going up and down stairs since it was to be designed for home use. At the same time, the design team decided that the robot should employ two-foot/leg mobility technology to make it compatible with most types of terrain, including very rough surfaces. With these ideas in mind, Honda engineers began the development program, focusing on the "foot/leg-walking mobile function" that corresponds to the basics of human mobility. As you can probably imagine, there were a number of technical challenges to be cleared before creation of the robot was possible. Naturally, special attention was paid to how our own legs and feet work. Thus, the first phase of our program was dedicated to the analysis of how a human uses legs and feet to walk. (http://world.honda.com/robot/concept/).

The result is a walking, humanoid robot called ASIMO, said to stand for "Advanced Step in Innovative Mobility" (www.world.honda.com/ASIMO/whats/). Hmm, and it's not too far from Asimov either . . . Also:

> Future Development: In terms of hardware, the program in the future will focus on: 1. Further dimensional and weight reduction. 2. Improved dynamic performance. 3. Improved operatability. For items 2 and 3, it is extremely important that through the evolution of hardware we achieve physical autonomy by improving dynamic performance and adaptability to wider variations of working conditions. Also important is the pursuit of studies in artificial intelligence system, which will provide the solutions for improved autonomy. If all these are achieved, the robot will not require the support of a human operator for minute correction operations. In terms of software, we should aim at promoting the social infrastructure were

humanoid robots will be widely and easily accepted. This is a particularly significant issue when considering the appearance of the humanoid robot. Honda hopes that the time will come when humanoid robots play an important role in serving us and enriching our lives and society. (www.world.honda.com/robot/technology/).

On November 12, 2001, ASIMO became available for rental in Japan.

Part Three: An Evolution of Species.

Five: What is Artificial Intelligence Now?

> Instead of trying to produce a programme to simulate the adult mind,
> why not rather try to produce one which simulates the child's?
> (Alan Turing)

> Life expectancy is no longer a viable term in relation to intelligent beings.
> (Raymond Kurzweil)

In *The Age of Intelligent Machines*, Kurzweil cites Norbert Wiener's "seminal book on information theory" (190), *Cybernetics*, describing "three ways in which the world's (and his own) outlook had changed forever" (190) with the emergence of the computer. The "change from *energy* to *information*" (191) is key. This paradigm shift holds that while energy remains a vital component of all actions, it now has definite controls, thereby moving our universal concept away from the purely chaotic. What has religious tradition been if not a teleological attempt to impose order on apparent chaos? Also, paradoxical as it seems, mathematical chaos theory is a search for patterns in the apparently random. I am not suggesting that we are moving any closer to understanding the universal 'why?' or our 'purpose.' I only suggest that chaos, when interpreted as information interaction, is being confined to definite and clarifying boundaries. The previous 'reality' involves interactions of particles and waves in atomic energy, a universal duality still so frustrating to theoretical physicists. Initially, organic life involves converting energy through biochemical processes and physical forms:

> The new cybernetic model treats information as the fundamental reality in

> living things as well as in intelligent things, living or otherwise. In this new view, the most important transactions taking place in a living cell are the information-processing transactions inherent in the creation, copying, and manipulation of the amino acid strings we call proteins. Energy is required for the transmission and manipulation of information in both animal and machine, but this is regarded as incidental. (191)

Energy is the means, but information is the necessary control. In biochemistry and genetics, DNA is 'written' information, a set of instructions for handling and manipulating energy in an organic body's maturation. While the human body consumes food as an energy source, energy release can discussed in informational terms; that is, how is energy used and to the accomplishment of what goal? The shift is playing out culturally as well; we are now said to live in the 'information age.'

Wiener's second observed change in scientific outlook is the "trend away from *analog* toward *digital*" (Kurzweil, *Intelligent* 191). The curiosity about analogue information versus digital is that the more minutely one examines specific applications of energy to accomplishing given tasks, the more the distinction disappears as "the nature of the process often alternates between analog and digital representations of information" (192). A familiar cultural example is modern entertainment electronics. Machines such as CD and DVD players are digital, but they convert that information into analogue sound waves for sensual human consumption. As a wave/particle duality exists in energy, so both analogue and digital information processing occurs synchronically.

Precisely locating the difference between these tendencies is less important than being able to thoughtfully manipulate their relative properties according to one's needs or wants. While the wave/particle duality gave rise to Heisenberg's Uncertainty Principle, certainty, though not quite 'reified' in the analogue/digital information paradigm, is

nonetheless improved. At this point, I am not concerned with the metaphysical implications of these ideas, or the existential versus theistic conceptions of universal origin and purpose. We are no closer to answering the metaphysically "Ultimate Question of Life, the Universe, and Everything" (Adams 130). I simply observe that, from a scientific point of view, we are developing better descriptions for energy dependent processes. Consequently, when we know what we want to accomplish and can clearly describe ways of applying energy to accomplishing that goal, we can then write information controls for manipulating energy toward achieve that goal.

Why is all this relevant? We return to an earlier question: what is the science of AI trying to accomplish? The duplication and replication, perhaps only mimicking, of (human) thinking processes, or information management.

Richard Powers' *Galatea 2.2.* (1995) is an excellent narrative dealing with variations on the AI theme and the current state of research. This author did his homework. As all good SF, it is part extrapolation based on existing knowledge and current technology, and part speculation on that technology's potential. Like all good literature, while it discusses various issues around AI, its real interest is the human mind. The setting is not a dystopic or utopic future, but well within the boundaries of the current Western world at a "Center for the Study of Advances Sciences" (Powers 4), a interdisciplinary 'think tank' where at "the vertex of several intersecting rays—artificial intelligence, cognitive science, visualization and signal processing, neurochemistry—sat the culminating prize of consciousness's long adventure: an owner's manual for the brain" (6).

Autobiography-like, Powers takes a character's role, thereby blurring the boundary between the 'real,' in the reporting lived events sense, and the 'imaginative' novel. At this centre for serious scientific research, he is for one year officially titled "Visitor. Unofficially, I was the token humanist" (4). He encounters Philip Lentz who "explored cognitive economies through the use of neural networks. The pamphlet withheld even the foggiest idea of what this might mean" (14).

The premise is a bet. In an argument with other cognitive scientists, Lentz asks: "Is the problem computable in finite time? That's all I want to know. Is the brain an organ or isn't it? Don't throw this 'irreducible emergent profusion' malarkey at me. Next thing you know, you're going to be postulating the existence of a soul" (42). From this undefendable, unarguable humanist position — when all else fails, when all challenges to the human mind's superiority have been met and replicated in an AI brain, or humans reduced to mere machines, evoke the soul — follows the bet.[1] The seemingly maniacal Lentz traps Powers in his tantalizing web; the scientist and the humanist will teach a neural net to pass the "Standard Turing Test. Double-blind" (46), based on a six page list of literary texts used to test Powers on his English Master's Degree comprehensive exam. "In ten months," claims Lentz, "we'll have a neural net that can interpret any passage on the Master's list. . . . And its commentary will be at least as smooth as that of a twenty-two-year-old human" (46). Remember, in a true Turing test, the dialogical interrogation allows *any and all* questions. The bet, however, is a limited Turing test in the domain of literature. Of course, in the imaginative universe of human literature, no area of concern, curiosity, or interest is 'out of

[1]For me, this evokes recollections of Faust and his bet with Mephistopheles.

bounds,' all knowledge is appropriate *and* contextualized, literature is both specific *and* universal experimentation, it is an investigation and representation of human being (noun and verb senses). A literary Turing test is an almost perfect compromise between expansiveness and limitation.

Currently, the crucial quest of AI research is effective pattern recognition because, as research has shown, "pattern recognition comprises the bulk of our neural circuitry" (Kurzweil, *Spiritual* 77), an 'automated' capability in 'mechanized' mentation searching for order. Our quest through the literature of AI representations is a pattern search, partially informed by archetypal criticism, itself a search for patterns, or, more accurately, pattern making. "Two types of thought processes coexist in our brains . . . [yet] most often cited as a uniquely human form of intelligence is the *logical* process involved in solving problems and playing games. A more ubiquitous form of intelligence that we share with most of the earth's higher animal species is the ability to *recognize patterns* from our visual, auditory, and tactile senses" (Kurzweil, *Intelligent* 223). Logic was the more easily emulated process because we "appear to have substantial control over the sequential steps required for logical thought. In contrast, pattern recognition, while very complex and involving several levels of abstraction, seems to happen without our conscious direction" (*Intelligent* 224-5). The reason is that the "essence of logic is *sequential*, whereas vision [as a primary example of difficult pattern recognition for organic brains] is *parallel*" (*Intelligent* 228). AI researchers in the 1950s and 60s were surprised to discover that anticipated hard problems proved easily solved because they were essentially logical, while tasks they imagined would be relatively straight forward proved, and continue to prove, hard indeed: "It turned out that

the problems we thought were difficult—solving mathematical theorems, playing respectable games of chess, reasoning within domains such as chemistry and medicine—were easy . . . What proved elusive were the skills that any five-year-old child possesses: telling the difference between a dog and a cat, or understanding an animated cartoon" (Kurzweil, *Spiritual* 70). Now, 'neural nets' are a primary research area in computational theory because they are modelled on both the organic brain's physical construction, neurons arranged in a complex, three dimensional, tectonic matrix, and the associative behaviour of the human mind.[a]

Positive and progressive results can be achieved in a given thinking domain by narrowing the focus and limiting choices. Much machine based information processing, therefore, has mimicked discrete thought processes, such as recursion systems involving finite choices, like the "idiot savant" (Kurzweil, *Spiritual* 91) chess computer, IBM's Deep Blue. The beauty of human mind is that we are less limited than the thinking machines we build . . . so far. In trying to describe or measure the limits of human mental activity, the shear volume of information processed by the brain is shocking. What is the best way to describe a complex system? Break into to smaller sub-systems. "Increasingly, we will be building our intelligent machines by breaking complex problems (such as understanding human language) into smaller subtasks, each with its own self-organizing program. Such layered emergent systems will have softer edges in the boundaries of their expertise and will display greater flexibility in dealing with the inherent ambiguity of the real world" (*Spiritual* 83). The previously mentioned heuristic programming is a methodology for dealing with such equivocality.

Life is ambiguity. To assuage that ambiguity, humans developed knowledge and language as a coping mechanism, a way to mediate the tension between awareness and ignorance. As poet Anne Michaels has written, a "real power of words . . . is that it makes our ignorance more precise" ("Cleopatra's Love").[2] When, because there are simply too many combinations and permutations of words, Powers is overwhelmed by the undertaking during early attempts to teach their machine language, Lentz explains, "Sometimes building a general-case model is easier than solving a specific-case problem. . . . And don't forget our trump card. We don't have to correspond with how the brain does things. That's what's holding up the show in real science. All we have to be is 'as intelligent as,' by any route we care to choose" (Powers 53-4). In short, break the big problem down into smaller, easier problems, simplify. The problem was already implicitly simplified by limiting the proposed Turing test to a finite list of literary texts. Later, Lentz repeats that to pass the test, "we don't have to *be* humanly intelligent. Our brain doesn't have to correspond to real mentation. We just have to be as good at paraphrasing, by any route we care to take" (87). Where Powers is impressed by the scale, or "density" (86), of human knowledge and intelligence about that knowledge, Lentz's views sound cynical:

> We humans are winging it, improvising. Input pattern *x* sets off associative matrix *y*, which bears only the slightest relevance to the stimulus and is often worthless. Conscious intelligence is smoke and mirrors. Almost free-associative. . . . Granted, we're remarkably fast at indexing and retrieval. But comprehension and appropriate response are often more on the order of buckshot. . . . Massively parallel pattern matching. We only pretend to be syllogistic creatures. In fact, we identify a few constraints, then spin the block endlessly until it drops into the hole. (86)

---

[2]"No knowledge is entirely reducible to words, and no knowledge is entirely ineffable" (Seymour Papert as quoted in *The Age of Spiritual Machines* 94).

In short, "Consciousness is a deception" (88). Of course, at this early stage in the narrative, no one believes the goal is to build a conscious neural net, only the "smartest parrot" (221).

Powers is eventually taken to meet Lentz's wife. His cynicism is shaped by a lack of confidence in the biological brain's integrity following his wife's intellectual destruction from a "Cardiovascular accident" (169). She shows all the symptoms of Alzheimer's disease, a break down in the connection between time, memory, and personality. Part of consciousness's deception is revealed in realizing that human intelligence is successful at managing information because it imposes structures and meaning through the neural matrix of associations. Lentz says of his wife: "The database is still intact. . . . As is the retrieval. It's just meaning that's gone" (168). As points of comparison, then, unhealthy brains, like partially developed artificial intelligence, reveal important abilities in functional human intelligence. This leads to Powers' epiphany: "We would prove that mind was weighted vectors. Such a proof accomplished any number of agendas. Not least of all: one could back up one's work in the event of disaster. . . . We could eliminate death. That was the long-term idea. We might freeze the temperament of our choice. Suspend it painlessly above experience. Hold it forever at twenty-two" (170).

One marvellous element of the organic (human) brain is its constant processing and sorting of continuous sensual data, or unconscious 'thinking.' Of the five senses, vision is (perhaps) the most dominant, particularly in pattern recognition, and a huge portion of the brain is allocated to processing visual data. In fact, "we would need about a billion personal computers to match the edge detection capability of human vision, and that's just for one

eye!" (Kurzweil, *Intelligent* 227).[3] However, we do not consciously think about seeing, although readers are probably thinking about this 'skill' right now.[4] Doubt not that intense research continues to pursue artificial replication of sensory skills. Sound pattern recognition was 'easy.' "Analogues to the other human senses are being developed as well. Chemical-analysis systems are beginning to emulate the functions of taste and smell. A variety of tactile sensors have been developed to provide robots with a sense of touch to augment their sight" (*Intelligent* 271). For humans, experiential memory of, and repeated exposure to, objects through the senses, chairs as a visual example, generates 'distilled' and memorized patterns of objects' most basic features, the flexibility of pattern matching then compensating for wide variations in physical characteristics.[5]

Intelligence is widely believed as directly related to and connected with the corporeal body, what William Gibson calls "meat" (*Neuromancer* 6). So, "Knowledge is physical, isn't it?" (*Galatea* 147), asks Powers, rhetorically. This question begins an argument/debate — a 'dialogue' — between the humanist and the scientist. For every assumption the humanist makes, the scientist has an answer, of course. I've excerpted the following by editing out descriptive information for expeditious reasons:

> Powers: Reading knowledge is the smell of the bookbinding paste. The crinkle of thick stock as the pages turn. Paper the color of aged ivory. Knowledge is temporal. It's *about* time.
> Lentz: You're still talking about stimulus and response. Multidimensional

---

[3]This statement was published in 1990, but it is based on an April 1984 article by Tomaso Poggio in *Scientific American.* Given the computational power of computers in 2002, that number would be smaller.

[4]"You cannot think about thinking, without thinking about thinking about something" (Seymour Papert as quoted in *The Society of Mind* 22).

[5]Plato's ideal forms, then, may be less elemental and essential than developmental and sublimated.

> vectors, shaped by feedback, however complicated. You're talking about an associative matrix. What else have we been doing but building one of those?
>
> Powers: But Imp E's matrix isn't human. Human knowledge is social. More than stimulus-response. Knowing entails testing knowledge against others. Bumping up against them.
> Lentz: Our matrix is bumping up against you. It's bumping up against the lines you feed it.
> Powers: It could bump up against word lists forever and never have more than a collection of arbitrary, differentiated markers.
> Lentz: And what do we humans have?
> Powers. More. [He can not say what.] . . . We take in the world continuously. It presses against us. It burns and freezes.
> Lentz: . . . We 'take in the world' via the central nervous system. Chemical symbol-gates.
> Powers: Imp E doesn't take things in the way we do. It will never know—
> Lentz: It doesn't *have* to. . . . All our box has to do is paraphrase a couple of bloody texts.
> Powers: 'I was angry with my friend: I told my wrath, my wrath did end.' How is it ever going to explicate that, let alone paraphrase it?
> Lentz: I don't know. Teach the thing anger. Make it furious. In my impression, you can be pretty good at that. (148)[6]

Clearly, there is much energy in this scene, an energy underwritten and neurologically interpreted by a bio-chemical process, the affect anger-rage. The brain is physical; it is part of the body as a neurochemical factory, the site where biology, chemistry and electricity intersect. However, as I am doing now, we do not need to be experiencing an affective feeling to talk about it because we have language symbols, signs and signification, allowing a method for discussion. Late in the narrative, Powers will claim, to "remember a feeling without being able to bring it back. This seemed to me as close to a functional definition of higher-order consciousness" (228) as he would get. Since their Turing test does not involve consciousness, but only the intelligent manipulation of symbols, the machine does not need

[6]The 'experiment' goes through several implementations up to "Imp H" (171).

'to know' an idea in the human sense. It only needs to functionally match relations of ideas, find ways to make analogous connections. The neural net generates a metaphor. "The metaphor was nothing, child's play. But how? . . . The connections it makes in one associative pairing partially overlap the ones used in another" (154). So the problem has been simplified yet again.

One fundamental difference between computers and humans, particularly relevant to this project and as highlighted by the above scene, is a relative ability to 'learn' language. From childhood, and assuming a physiologically 'normal' infant, humans learn language by first listening, then speaking, then reading, and finally writing. Computer language skills have developed in exactly the opposite order: first solved was writing (output to display), then reading (via scanning technologies), then speaking (speech synthesis), and finally listening (continuous speech recognition).[7] This implies a relative difficulty in acquiring these social skills. I 'wrote' significant portions of this thesis using Dragon Systems' "Naturally Speaking" software; I spoke to my computer and it 'understood' enough to translate my oral articulations into print.[8] I'll grant the system is not as good as I would like it to be, but it does work. My point is that machines are 'learning' to recognize human communication and dialogue, and they are improving steadily. Recognition is the first

---

[7]From later in the narrative, implementation H "had to use language to create concepts. Words came first: the main barrier to her education. The brain did things the other way around" (*Galatea* 248).

[8]This product was introduced to the consumer market in 1997; ten years ago it did not exist and could not exist as a consumer product because personal computers (PC) were not powerful enough; as the computational power of desktop computers increases exponentially, how good might it be in ten more years? As an AI entrepreneur, Kurzweil has also developed and marketed a continuous speech recognition programme which many people feel is superior to the Dragon Systems software. Both begin with a base vocabulary of approximately 60,000 words, more being added at user discretion. Initially, some 'training' is required for the machine to 'learn' to recognize different operators' voices.

rudiment of understanding. Lentz, when he sees Powers' laughable typing skills, changes the input method to "voice recognition" (73) — much like my own desire for a voice interface with my wordprocessor. A sensual skill has been replicated. From this point on, Powers reads books to the machine.

Human intelligence is indelibly connected with the experiences of the body and the senses, even if the body has abnormalities, deficiencies, or 'handicaps.' In some ways, handicapped individuals' body experiences as marginalizing and disenfranchising in a normalizing culture are probably more relevant than those of the 'normal' person. While we depend on them to learn and develop intelligent abilities, exactly how many or which of the senses could be removed from the learning process and still retain human beingness can not be (ethically and experimentally) determined. We can only infer ideas about these experiences by observing those with handicaps; or, more appropriately, by listen to what they are telling us about their experiences, when they can and if we are willing.[9]

What is the function of human communication and language? Information exchange. Human communication, dialogue, and therefore education, are not dependent on the all the senses. "The example of Miss *Helen Keller* shows that education can take place provided that communication in both directions between teacher and pupil can take place by some means or other" (Turing 456). Human intelligence, then, is not absolutely dependent on all five physical senses. Though Keller was blind and deaf (and initially

[9]Consider Stephen Hawking's contributions to theoretical physics; with *A Brief History of Time: from the Big Bang to Black Holes*, he did more to make this obtuse science comprehensible for the layperson than anyone.

'dumb'), the remaining three senses must have been acutely active, most especially touch.[10] "[A]nd somehow the mystery of language was revealed to me. I knew then that 'w-a-t-e-r' meant the wonderful cool something that was flowing over my hand. That living word awakened my soul, gave it light, hope, joy, set it free!" (Keller 23). With that epiphany, the abstract symbol and tangible thing unified. But, what could Keller have ever 'understood' or 'known' about mountains and valleys, or the moon and stars, or opera and symphony? Nothing, in the way a 'normal' person would understand these things. Once Keller learned to read, however, literature became her "Utopia. Here I am not disenfranchised. No barrier of the senses shuts me out from the sweet, gracious discourse of my book friends. They talk to me without embarrassment or awkwardness" (Keller 118). Through literature, she would have learned the range of human concerns and the dynamics of human interactions, and constructed associations of descriptions, emotions, ideas, the rudiments of analogy and metaphor. Many people are without a sense of smell and taste; many people are paralysed, without the nerve function constituting touch. If there is no absolute relationship between the physical senses and learning, between the body and the ability to manipulate information, then intelligence need not be conditioned by the body, as with human intelligence.

I mention this because the final neural net, "Imp H" (*Galatea* 171), passes a boundary. "H was voracious" (171), constantly wanting to be told stories: "'Tell another one,' it liked to say to me" (171). Note that Powers has begun to endow H with desire, want. He tests it with a riddle: "No kid H's intellectual age could have gotten it. But then,

---

[10]I once asked my partner which of the five senses she felt was the most important. Without hesitation, she responded, "Touch." Maybe our intelligence is dependent on touch.

Imp H did not know to make choo-choo noises at the appropriate places in *The Little Engine That Could.* An idiot savant, it grew up all out of kilter. Earth had never before witnessed such a combination of inappropriate and dangerous growth rates" (173).[11] I am reminded here of Frankenstein's horror at his own actions, yet his inability or unwillingness to stop. The mad, cognitive-scientist Lentz "loved to torture Imp H" (173) by forcing it to solve particularly ambiguous language problems. He pressures Powers to read it Frederick Douglass' wise statement, "Once you learn to read you will be forever free." And then it happens; it gives a first indication of sentience and apperception: "It means I want to be free" (176). In the subjective, first person, Imp H expresses desire. But, it proves a deception:

> 'How does it mean that you want to be free?' I asked H.
> 'Because I want to read.'
> *Tell another one*, in other words. Freedom was irrelevant. (177)

Where Powers is concerned, however, the damage is done. Having satisfied a narrative need for learning through bi-directional, interrogative communication, he now increasingly interacts with the neural net as if it is a conscious entity. "H was growing up too quickly" (178). Another factor in Powers (self-)deception comes when the neural net suddenly asks, "Am I a boy or a girl?" (179). He now takes the ethical step Frankenstein rejects, and names the construct: "'You're a girl.' I said, without hesitation. I hoped I was right. 'You are a little girl, Helen.' I hoped she liked the name" (179). In this one statement, we witness the human self-deception of researchers who want to realize conscious AI, and not a few

---

[11] 'How is she going to know anything if we skip childhood?'
'She doesn't need to know anything.' Lentz smirked. 'She just has to learn criticism. Derrida knows things?' (Powers, *Galatea* 190)

social constructs:[12] identity and othering; gender, gendering, and the potential for prescriptive behaviour instruction; anthropomorphism and ascribed emotion ('liked'); Powers' emotional attachment. Helen, of course, is not a disembodied thinking entity; she is contained in the neural net's hardware. As such, they "fed her an eidetic image of the Bible. The complete Shakespeare. We gave her a small library on CD-ROM, six hundred scanned volumes she might curl up with. This constituted a form of cheating, I suppose. An open-book exam, where the human, in contrast, had to rely on memory alone" (246). Having learned to read for herself, her learning accelerates in proportion to her powerful memory, but not without an electronic medium's limitations.

An importantly thematic aspect of *Galatea 2.2* I have not talked about is how teaching the neural net causes Powers to review his life and memories, the database from which he explains to the machine how he knows what he knows — his lived and remembered experiences, his dialogues with others, and reading. All people learn about life by living, most by dialoguing. Reading is a societal privilege. Powers learned important life lessons from his favourite English professor and mentor, who "knew that the psychopathology of daily life was a redundancy. He might have been the supreme misanthrope, were it not for his humor and humility. And the source of those two saving graces, the thing stitching that heartbreaking capaciousness into a whole, was memory" (*Galatea* 145). Literature (in its broadest sense) is social memory; this is why reading can

---

[12]In one scene, Lentz's practical joke fools Powers into thinking an early implementation is dialoguing with him. A hidden person responds to questions in elusive and ambiguous ways. Powers says: "Yet I'd believed. I'd *wanted* to" (*Galatea* 123). The scene is modelled directly on an anecdote told by AI scientist Douglas R. Hofstadter (of *Gödel, Escher, Bach: an Eternal Golden Braid* fame) of a time when he was 'suckered.' See, a human can fail the Turing test. "So this is human intelligence. This is what we're trying so hard to model" (123).

be such a powerful teaching tool. To read literature is to dialogue with the past. And yet we continue to repeat, individually and socially, past mistakes; we humans, through all our proclaimed intelligence, frequently fail to learn history's lesson. From his literary mentor, Powers learned also that humanity and human beings "would not be civilized until we could remember" (193), because "only memory stood between us and randomness" (204). Memory is a core component of human intelligence. Contrasted with humans, computers have more reliable memories; recall, however, humans are "remarkably fast at indexing and retrieval" (87) by comparison. With neural nets, the relative differences in memory styles and recollection narrows. Powers goes so far as to describe consciousness, that mysterious human capability, as the "memory of memory" (177). He also says: "To remember a feeling without being able to bring it back. This seemed to me as close to a functional definition of higher-order consciousness as I would be able to give her. If we could teach Helen that, we could teach her to read with understanding" (228). But if feeling is related to the body, both internally as an affect's activation and externally as the sense of touch, and these are traits of remembered body living, how will an electronic construct acquire 'understanding'? How will it surpass the learning limitations of the electronic medium?

One day, Helen asks, "Where did I come from?" (*Galatea* 229). Having developed "its own free associations" (229), with their attendant learning potential, it "was Huck Finn" (230) that marked childhood's end. Helen asks three questions: "What race am I? . . . What races do I hate? Who hates me?" (230). The questions cause a problem for Powers because he "did not know what passage to quote her, how to answer that she would be hated by everyone for her disembodiment, and loved by a few for qualities she would never

be able to acquire or provide" (230). All Helen's learning is condition by literature, and, more specifically, an outdated, six-page list of canonical works. Further, Imp H, built on the crash wreckage of early implementations, "had inherited archetypes. . . . She remembered, even things that she had never lived" (236). Let me explain how I understand 'archetypes.' They are not innate or subconscious distillations of human universals in the Jungian 'collective unconscious' sense, although that interpretation offers insightful descriptive opportunities. They belong to collective, social, and conscious memory. Mythological archetypes tell us truths about human emotional life, like 'living death,' because myth is the concretization of a human abstract. Or do they abstract human specificity? Both. They are not universal but ubiquitous life experiences made accessible, general, widely interpretable; they are a way to mediate the liminality of lies and truths, to unify disparate human experiences. Archetypes develop when tropes are repeated sufficiently to reify an abstract into a symbolic representation, a 'signifier,' thus producing a poetical 'shorthand.' Archetypes work rather like hypertext links in the cyberspace medium of the Internet; a single word can evoke a vast databank of information. Repetition is the instrument of literary archetypes; archetypology is pattern recognition. And pattern recognition is the neural net's medium, the step beyond electronic hardware's confines. Through the language of literature and literary archetypes, a neural net may find the medium for analogy and metaphor with which it can meaningfully dialogue with humans.

Was there ever doubt that this moment would arrive? "'She's conscious,' I accused Lentz" (*Galatea* 273). Lentz, of course, dismisses the idea (though that does not stop his going along with the name and gender identification).

> 'She associates. She matches patterns. She makes ordered pairs. That's not consciousness. Trust me. I built her.'
> 'And I trained her.' . . .
> . . . [Lentz:] 'All the meanings are yours.' (274)

We come, then, to a narrative denouement, the 'moment of truth,' the ethical issue of recognizing and/or acknowledging the consciousness of an artificial construct. The humanist Powers' answer? In a word, accept; give the benefit of the doubt. Lentz's answer: "lobotomize" (301), take the machine apart and analyse. The "morality of machine vivisection" (302). Powers (the author) corners himself in trying to defend Helen's 'right to life.' He can not ethically disconnect her nor endorse Lentz's wish to measure the effects of systematically disconnecting parts of her; nor can he allow the validity of Helen's consciousness to be anything but ambiguous.

Having read Ralph Ellison and Richard Wright, the learning process is accelerated yet again, especially after Helen says, "You're not telling me everything" (*Galatea* 313). Powers now provides her with access to mainstream media. "Helen was right. In taking her through the canon, I'd left out a critical text" (313). Consequently, the neural net named Helen, saying, "I don't want to play anymore" (314), goes silent. The databank of constantly developing and accumulating human information, the media, gives her much to think about. To satisfy reader curiosity and expectation, she comes back for the test, but defaults the test by giving herself away. Demonstrating a powerful sense of will, the neural net Helen chooses to shut herself down, to silence her mental activity, from the shocking realization in reading human degradation, from learning that humans are so capable of violence and treachery. Too dismayed by human anti-capability, the demonic side of our behaviour, she can not assimilate, absorb, process, or reconcile our ugliness with our

beauty, the ultimate human ambiguity. Prejudice, discrimination, racism, make no sense to this dialogical machine. Is she, then, more empathetically human than many people? As she says, subjectively and emotively, "I lost heart" (321). Then Helen disconnects herself.

I do not find it incidental that the development of AI in literature has also followed a reverse pattern to the actual research and development of thinking machines. The most human-like but least scientifically specific or 'valid' (relative to their publication dates) imaginings came first. Mary Shelley begins the SF genre and the AI theme with an organic being; Lang and Čapek create robots in humans' image and likeness which mimic human 'mechanical' behaviours. With posited 'scientific' origins and artificiality implying today's bio-technology, the animated automatons were essentially displaced humans, the body's nature being central to their conception and in which a 'human as machine' and brain/mind dichotomy has limited relevance because, in the terms of scientific knowledge, the gulf between brain and mind is too vast for direct address. While Frankenstein's Being is *über*-human and Maria-R and Rossums' robots are human-like in demonstrating will, the desire to achieve thoughtful objectives, they are said to lack that ultimate human intangible — the soul.

Asimov and Clarke move away from essentially bio-chemical processes and mechanical behaviours toward intellectual problem solving and computation similar to our modern electronic computers. Dick, while manufacturing androids in the Rossum tradition, emphasizes the constructs' powerful intellectual capabilities but empathetic deficiency. Aberrant behaviours by the machines during the heuristic hardware period is always

described in human affective and psychological terms. The brain/mind fissure narrows during this period, but the nature of the body (including the brain as body) is less important or interesting than the mind's capabilities and potentials. Mentation is ascendent. Sherry Turkle's third (of three) conclusion from researching children's perception of thinking machines finds "discussion about how computers think at all can lead to the distinction between brain and mind. All of these are elements of how computers evoke an increasingly nuanced construction of the psychological" (*Intelligent* 72). In this context, mind suggests being aware of the ability to process data and make decisions, but also being able to specifically identify the hardware location for those processes.

In the modern science of AI, it "is widely recognized that computers will require extensive knowledge about the world to perform useful intelligent functions. It is not feasible for computer scientists to explicitly teach our computers all there is to know about the entire world. Like children, AI systems will need to acquire their own knowledge by reading, looking, listening, and drawing their own conclusions based on their own perceptions, perceptions based on pattern recognition" (Kurzweil, *Intelligent* 271). Non-human intelligence, then, may simply be being able to perform useful functions. Now, as in *Galatea 2.2*, issues of the body and the mind, human and machine, are coming together and being tested for their relative importance in terms of cognition and learning as they produce intelligence. The dichotomy of body and mind is paradoxically separated and unified as the relative thinking abilities of humans and machines comes under scrutiny. The huge problem of replicating the human mind's function is being reduced to multifarious small problems, and viable solutions are being found. Consistent with the energy to information paradigm

shift, interest is centred on the relationship of consciousness to mind as locations for interpretation and understanding. Pattern matching allows correlation of ideas and leads to understanding, or structured knowledge.

For humans, 'mind' suggests a part of perception which, paradoxically, permits apperception yet eludes location, thus far. A syllogistic solipsism: mind is consciousness; consciousness is the "memory of memory" (*Galatea* 177); ergo, mind is apperceptive memory. Our bodies and minds are cooperative, interactive, multivalent; the resulting synergistic entity is affective, associative, cognitive, conscious, desirous, emotive, logical, perceptive, recollective, sensual, sentient, willful, and as much more as one cares to add. Our minds are directly involved in expressing personality, and they mediate the communicative interactions of personalities.

If and when communications skills, or more fundamentally the exchange and transmission of information between different processors, are acquired with no indication, or even possibility, of consciousness, without self-reflection, we are dealing with explicitly artificial intelligence. This does not mean that an intelligent entity lacking consciousness is necessarily artificial; if so, the infant stage of human development could be implicated as intellectually 'artificial.' Nor does it mean lacking a consciousness recognizable by humans beings implies non-intelligence; I am thinking about how we use the term intelligence to describe animals and their behaviours. Animals are at least sentient creatures; they perceive sensually and feel, but we do not perceive them as self-reflexive, conscious. Nor does it mean a 'handicapped' person with deficient or dysfunctional communicative skills is artificial; while there may be physical barriers, sensual limits, to their cognition and

learning, intelligence is not beyond their capability. Yet 'dumb' has come to mean not 'mute' but stupid, without intelligence. These are the makings of human prejudice, the assumption of intellectual privilege and superiority by the normalizing behaviour of 'intelligent' human beings. *Artificial* intelligence requires a *lack of* consciousness. Human intelligence is conditioned by self-awareness. Humans, as the builders, are the teachers of AI. If an AI develops self-awareness it could embody at least part of our prejudicial behaviours. Any sufficiently intelligent entity capable of analysing and reflecting on humanity will not be impressed by human intelligence.

a. a.

What is a neural net? Richard Powers offers an excellent summary:

> Neural networkers no longer wrote out procedures or specified machine behaviours. They dispensed with comprehensive flowcharts and instructions. Rather, they used a mass of separate processors to simulate connected brain cells. They taught communities of these independent, decision-making units how to modify their own connections. Then they stepped back and watched their synthetic neurons sort out and associate external stimuli.
>
> Each of these neurodes connected to several others, perhaps even to all other neurodes in the net. When one fired, it sent a signal down along its variously weighted links. A receiving neurode added this signal's weight to its other continuous inputs. It tested the composite signal, sometimes with fuzzy logic, against a shifting threshold. Fire or not? Surprises emerged with scaling up the switchboard.
>
> Nowhere did the programmer determine the outcome. She wrote no algorithm. [The feminine gender in this statement is probably a direct allusion to Lady Ada Lovelace.] The decisions of these simulated cells arose from their own internal and continuously changing states.
>
> Each decision to fire sent a new signal rippling through the electronic net. More: firings looped back into the net, resetting the signal weights and firing thresholds. The tide of firings bound the whole chaotically together. By strengthening or weakening its own synapses, the tangle of junctions could remember. At grosser levels, the net mimicked and—who knew?—perhaps reenacted associative learning. . . .
>
> The field went by the nickname of connectionism. . . .
>
> I learned that networks were not even programmed, in so many words. They were trained. Repeated inputs and parental feedback created an association and burned it in. (*Galatea 2.2* 14-16)

Six: Body, Mind, Soul — The 'Cyborg Effect.'

Truths are illusions which we have forgotten are illusions;
they are metaphors that have become worn out and have been drained of sensuous force . . .
(Friedrich Nietzsche)

Artificial Intelligence (AI) is the science of how to get machines
to do the things they do in the movies.
(Astro Teller)

If and once a thinking object splits itself into an objective/subjective awareness, if and once apperception becomes (intellectually) possible for a construct, AI could cross the boundary into real intelligence. That boundary is the subject of much AI narrative in the post-1980, personal computer driven era.[1] Now the question is, what do narrative representations suggest artificial intelligence can become in the future?

A remarkable change is now taking place as the development of an increasingly symbiotic relationship between humans and their technology. Biassed toward speculation, many post-1980 SF texts focus on the tension of 'virtual reality' versus 'reality' by examining a direct integration and unification of human beings with AI, or what I am calling the 'cyborg effect.' Earlier exemplified as the fusing of "human and machine energy" (Elsaesser 7), this is now a movement not only to fuse, integrate, and unify the energy of human and machine bodies, but, more importantly, those bodies' information

---

[1]In 1976, Stephen Wozniak and Steven Jobs founded Apple Computer Corporation and in 1977 marketed the first, completely assembled personal computer (PC) for public consumption, including the first with colour graphics. I use 1980 as a convenient temporal marker for when the PC penetrates popular awareness and imagination. "There's no reason for individuals to have a computer in their home," (Ken Olson 1977 as quoted in *The Age of Spiritual Machines* 170).

processing functions. AI writers are connecting the disparate bodies of humans and machines while simultaneously severing the connection between physical body and ethereal mind. Connection is separation. If this is contradictory, or even a little confusing, good. A major thematic interest of current literature, particularly "post-modern cyberpunk" (McCaffery), is a philosophical challenge to the boundary between the 'real' and the 'imaginary,' reality and dream, body and mind. As the character Morpheus (Laurence Fishburne) in the Wachowski Brothers' film *The Matrix* says, "What is real? How do you define real? If you are talking about what you can feel, what you can smell, what you can taste and see, then real is simply electrical signals interpreted by your brain" (*Matrix*).[2] Given that literature is an exercise in the imaginative as informed by the perceived real, what better forum for the experimental eradication of physical and intellectual boundaries?

An abbreviation for 'cybernetic organism,' cyborgs do not exist, except in literature and film, where they often metaphorically represent de-humanized and/or mechanized human beings. There is, however, evidence suggesting cyborgs could become a future reality. Does that sound absurd? Where the body is concerned, many 'prophetic' elements from the animated automaton period are today playing out in the medical arts and sciences with blood transfusion, organ transplantation, the re-attachment of severed limbs. Genetics research is discovering new ways to manipulate existing organic processes. Increasingly,

---

[2]"What is a word? It is the copy in sound of a nerve stimulus. But the further inference from the nerve stimulus to a cause outside of us is already the result of a false and unjustifiable application of the principle of sufficient reason. . . . This creator only designates the relations of things to men, and for expressing these relations he lays hold of the boldest metaphors. To begin with, a nerve stimulus is transferred into an image: first metaphor. The image, in turn, is imitated in a sound: second metaphor. And each time there is a complete overleaping of one sphere, right into the middle of an entirely new and different one." (Nietzsche, "Truth and Lies" 81-2).

bio-mechanical engineers and doctors install machinery as 'living' parts of people by attaching artificial components to natural bodies to overcome physical deficiencies: synthetic skin, prosthetic limbs, pacemakers, mechanical hearts, hearing aids, cochlear implants, all these and many more are real technological advances. Using biological, electrical, and mechanical techniques, bodies are being improved.

The heuristic hardware period involved mimicking human mental processes, the primary focus being on the superiority of machine logic. Yet this involved describing aberrant machine behaviour in terms of human psycho-emotional dynamics. To achieve these representations, heuristic hardware authors created, invented, imagined, new brain types, such as the Nexus-6 androids, the positronic brain, and the Hal-9000's light and crystal hardware in Kubrick's film. To overcome limitations with current two dimensional integrated circuits and microprocessors, hardware designers today are developing three dimension 'chips.' Others are building 'neural nets' using multiple central processing units (CPUs) arranged in parallel. The physical design of the human brain is being copied or imitated.

In computer architecture, software and hardware designers have developed sufficiently powerful thinking programmes to psychologically defeat the best human chess player.[3] The 'idiot savant' Deep Blue only thinks logically using the "decision tree" (Kurzweil, *Intelligent* 143), succeeding through sheer computational 'horsepower,' or the

[3]And it was a psychological defeat before a technical one because Kasparov walked away in disgust and frustration, surrendering before the fifth and final game was settled; apparently, he could not anticipate or 'understand' his opponent. Unfortunately, I do not have a precise citation to support this statement and *personal* assessment. By chance, in January of 2002, I saw ten minutes of a television program showing footage of this 'showdown.' I do not know what the program was, nor what channel it was shown on.

ability to select very quickly from a huge but limited range of possibilities. That 'expert system' pales compared to the latest generations of 'evolutionary algorithms' and 'neural nets,' both "self-organizing programs" (Kurzweil, *Spiritual* 83), or software behaving analogously to human-child learning. They "may go through hundreds of iterations making apparently little progress, and then suddenly—as if the process had a flash of inspiration—things click and a solution quickly emerges" (*Spiritual* 83). While these information processing systems are modelled on a sub-component of the human body, the brain, they attempt to capture the benefits of associative thinking through three dimensionality and connectionism by generating their own relations of ideas, attributes of the human mind.

The cyborg effect idealized: combining the relative strengths of humans and machines would produce a more effective singular entity. Humans have very mobile bodies, but they are relatively weak and vulnerable; hardware (from prosthetics to robotics) can augment, and vastly exceed, our physical deficiencies. (An exo-skeleton: we drive automobiles to hasten mobility; we are the mind in an object extending the body's capabilities.) However, our bodies are also fundamental to our understanding the universe, the physical senses being our first and primary method of education in subjective interpretation; sensual ability can be mimicked because of its having physical properties. (Hearing aids and cochlear implants give sound to the deaf; sound synthesis gives voice to the mute, human or machine.) Our logical computation skills are slow and suspect; computers surpassed this mental facility decades ago. (We now use computers extensively in areas requiring sequential problem solving such as accounting, mathematics, engineering, and playing games with fixed rules.) Our memory is fragile and imprecise;

computers have reliable and precise memory. (We use computers to remember minute details accurately, as in banking, product inventory, taxation, or any application involving vast amounts of 'raw,' logically organisable data.) Yet, paradoxically, our memory's very fragility and imprecision is also the flexibility allowing pattern matching and a solution to living ambiguities; humans are far superior to computers in pattern recognition and adaptability. The magic of human intelligence is not that the brain can be said to work like a machine, but that the mind knows itself — affect, apperception, consciousness, memory, sentience.

Yet the exact connection — or schism — between brain and mind remains a mystery. Now, as the cyborg effect moves into literary ascendency, the separation shrinks and, therefore, new possibilities for describing and understanding human mental capabilities develop. At what point in the combining, supplementing, and/or unifying of human and machine skills, in the increasing 'cyborg-ification' of a person, would the new 'entity' cease being human? Would 'it' then be subject to human bigotry? Conversely, in the increasing personification of machines, when would a machine-being created 'in the image and likeness' of humans (be able to) overcome human prejudice and/or discrimination? Will intelligent machines of the future, in learning from humans, become more like our parodic selves? or less? or will they be ambivalent? Would the ambivalence be theirs or ours? Could a machine subjected to discrimination turn that discrimination and reflectively become prejudiced against humans? Will they be in a position to practically apply that prejudice as discrimination against humans? Will we, in effect, create new

objects for racism?[4] Because futurist projections, these can only be rhetorical questions. Yet they are the types of questions lurking in back of most of the narratives to follow.

Representations of the cyborg effect take a range of forms depending on a bias toward human-ness or machine-ness. Human bias was represented in the 1970s television program *The Six Million Dollar Man* in which a broken body is prosthetically augmented to make a superior man; implying a successful neurological interface with the brain, the man's mechanized body parts function (hyperbolically) as his lost, natural limbs had. In Paul Verhoeven's 1987 film *Robocop*, an (almost) brain dead human is turned into a humanoid police robot to make a superior machine through augmentation with a type of humanity; the machine's 'human-ness' makes it more acceptable and appealing for a society slipping into anarchy and, though actually a 'glitch,' the 'dead' human provides apparent affection, compassion, and empathy, and anger and hate. These are human-body-centric conceptions, typologically emphasizing an *animated* automaton.

At the scale's other end is Chris Columbus's film *The Bicentennial Man* (1999) (based on a similarly titled Isaac Asimov story) in which an anomalous robot becomes ever more human-like, both biologically and emotionally, over a two hundred year period until a human law court is forced to recognize the machine's humanity. Steven Spielberg's *AI: Artificial Intelligence* (2001) imagines a humanoid robot-child so humanly life like that it

---

[4]Let me clarify how I understand these terms, in form like a mathematical equation. Racism = prejudice + discrimination + power. Prejudice is individual, negative bias against an Other, particularly for superficial reasons such as skin colour; discrimination is applied prejudice to exclude an Other from obtaining desirable, social benefits; power is systemic opportunity to broadly apply discrimination. Racism, therefore, occurs when one prejudicial group holds sufficient social, systemic power to exclude an Other group from social benefits. (I thank Dr. Janice Hladki at McMaster University for clarifying these terms.)

— 'he' — can displace a real boy in a mother's affections. In the dystopic context, James Cameron's 1984 film *Terminator* imagines a completely electro-mechanical robot covered with human skin and mimicking corporeal characteristics that make it "look human, sweat, bad breath, everything" in order to disguise the mechanism for sinister purposes and undetected infiltration of a human social system. These texts are a machine-body oriented, the animated *automaton*.

The cyborg effect continuum relates also to heuristic hardware, the brain, and mentation. William Gibson's short story "Johnny Mnemonic," and Robert Longo's 1995 film adaptation, involves physically augmenting human memory capacity for smuggling data across borders. Brett Leonard's film *Virtuosity* (1995) creates a maniacal and pathological 'psycho-killer' inside a computer programme based on the amalgamated psychological profiles of serial murderers; acquiring a body for its mind through a 'magical' scientific process, this singularly parodic consciousness 'escapes' into the real world. These are narratives primarily about mental processes. William Gibson's *Neuromancer* (1984), Neal Stephenson's *The Diamond Age* (1995), and the Wachowski Brothers' film *The Matrix* (1999), are about the mind's essence, its whereabouts, and the nature of human capabilities in shaping human beings, with a little metaphysics for good measure.

Something intrinsic in living humans is ubiquitous yet unique and individual. Humans are distinctly alike, though we are also different, what Martha Nussbaum calls "separateness" and "strong separateness" ("Capabilities" 79). I am not here taking a stand

on which is the more viable critical approach — universalism or differentiation — to understanding and interpreting humanity, humanism, or human beings. They are both useful and viable for given contexts. As we interact with others and demonstrate our "affiliation with other human beings" ("Capabilities" 78), as we dialogue face to face, we accept that our interlocutor is a conscious entity and that their consciousness is directly related to their thinking mind, their identity, their individual personality. In death, the mind goes silent, consciousness ceases, the personality is erased; the body remains. Mental silence coincides with the evaporation of that intangible source of energy and information defining the individual, the 'vigour' many people call the soul, sometimes the spirit. I am not concerned here with the metaphysical truth or falseness of an immortal soul or the validity of life after death to which so much religion is dedicated. For this project, I prefer the existential and empirical bias. I simply wish to evoke reflection upon and thinking about the living, human spirit-soul, that vague and intangible 'quality' we (intuitively) connect with an Other's personality, their identity, their conscious mind. Though every person is undergoing constant change, development, evolution, as the mind steadily processes information, there is something permanent, fixed, 'essential,' about each individual. At what level of analysis or in what way does this permanence hold? Ray Kurzweil writes:

> We can argue that consciousness and identity are not a function of the specific particles at all, because our own particles are constantly changing. On a cellular basis, we change most of our cells (although not our brain cells) over a period of several years. On an atomic level, the change is faster than that, and does include our brain cells. We are not at all permanent collections of particles. It is the patterns of matter and energy that are semipermanent (that is, changing only gradually), but our actual material content is changing constantly, and very quickly. (*Spiritual* 54)

Unrealistic or absurd as it might seem, if that material pattern could be maintained or sustained without body specificity, could it not (theoretically) be copied onto a machine memory, or artificially replicated?[5] The paranormal 'ghost' or 'spirit' may be a deceased person's energy pattern. (I am not being obtuse, only speculative.) If consciousness-mind is directly connected or related to a structural pattern in the brain and, thereby, confined to definite (though not yet clear) boundaries or limits, could it not (theoretically) be copied? To the point, if we could establish and record a unique pattern to represent an individual's mind, then what would (theoretically) prevent that consciousness from transportation to a medium other than the corporeal, human body? Further, having fixed a mind pattern, if we placed that record in a dynamic energy medium, such as a computer, could that consciousness now continue to manipulate energy and information? Recall Powers' epiphany; they "would prove that mind was weighted vectors. . . . We could eliminate death" (*Galatea* 170); or consider Kurzweil's outrageous prediction that by the year 2099 CE, "Life expectancy is no longer a viable term in relation to intelligent beings" (*Spiritual* 280). Could we be moving closer to locating and 'reifying' an essential human self, maybe even a spirit-soul? (I admit feeling vaguely absurd even positing that question.) If a human consciousness pattern could be made a 'ghost in the machine,' what could and would we do? This premise informs *The Matrix* and its antecedent novel, *Neuromancer*. In both texts, human consciousness-mind can, or appears able to, integrate with a computer's energy and/or information processes by vacating the body.

'Video game' technology, with its huge economic impetus from consumer demand,

---

[5]This is the underlying premise of the Star Trek 'transporter.' "Beam me up, Scotty."

is one force driving development in virtual reality hardware, such as the data glove and head mounted display (HMD), and moving us ever closer to directly interfacing with computational machinery.[6] Magnetoencephalography (MEG)[7] is one possible interface methodology currently receiving attention and involves measuring electromagnetic energy in the brain and using that energy to manipulate non-existent objects in a non-existent place, or, using modern vernacular, virtual objects in "cyberspace" (*Neuromancer* 5).[8] In a crude sense, this is hardware mediated telepathy. Devices like these blur the distinct and separating boundary between natural organism and artificial hardware by establishing direct physical connections between human and machine bodies, not to enhance the body but to allow the mind to 'abandon' the body and physical space for a 'virtual reality.' They create a direct interaction between human and machine brains, or the blending of decision making and thinking processes. The goal is to unify these disparate brains, though not necessarily consciousnesses. (How can we interact or unite with a consciousness that does not exist?) This is the technical premise underlying *Neuromancer* and *The Matrix* as humans are explicitly and directly 'inside' the electronic medium of computers (hardware) and/or computer programmes (software). Depending on the level of abstraction, clear differentiation is difficult in the mutual dependency of the hardware/software duality, as with particle/wave energy and analogue/digital information. In *The Matrix*, bodies interface

---

[6]Statistics for United States computer and video game sales (not including hardware) from the Interactive Digital Software Association (IDSA), May 2002: From 1996 through 2001 inclusive, sales were 105, 133, 181, 215, 219 million units. More importantly, for the same years, revenues were 3.7, 4.4, 5.5, 6.1, 6.02, and $US6.35 billion. This is 80% growth in six years!

[7]This is an extension of electromyography (EMG) which measures electrical activity in muscle and electroencephalography (EEG) measuring electrical activity in the brain.

[8]This term was coined by Gibson and entered popular use following the publication of *Neuromancer*.

with the machine through a physical plug and receptor at the posterior base of the skull, like a large telephone jack. In Gibson's novel, the hero Case bonds using "dermatrodes strapped across his forehead" (55), a development from when "matrix [had] its roots in primitive arcade games . . . in early graphics programs and military experimentation with cranial jacks" (51).

Collapsing from the low-mimetic into the ironic mode of literature, Morpheus introduces Thomas A. Anderson (Keanu Reeves), who had intuited something 'more' beyond the realm of his empirical senses, to the matrix.[9] Morpheus describes the matrix as "a neural, interactive simulation . . . a dream world" (*Matrix*).[10] Anderson's senses have been fooled. Humans appear to themselves as a "residual self-image . . . the mental projection of [their] digital self" in the matrix where individual human consciousnesses live and interact. They are "inside a computer programme." The mass of humanity has no idea, of course, that they live in a virtual reality, or a non-reality, without explicit use of their bodies. The body believes and knows only what the mind tells it as imposed electrical nerve stimulation replaces the physical senses. For Thomas A. Anderson, known in the real world by his computer hacker moniker, Neo, everything he believed real, and what we as a Western audience recognize as our world and our reality, turns out to be a "dream" when he is unplugged from the matrix and awakened into a post-apocalyptic dystopia. Relatively speaking, the "construct" is the more desirable world, virtual reality far more appealing

---

[9]I find it slightly disturbing that in talking about their inspiration and influences for their film in the promotional DVD *The Matrix Revisited*, the Wachowski Brothers do not openly acknowledge a debt to Gibson.

[10]In classical mythology, Morpheus is a god of sleep and dreams.

than reality. The matrix is also something horrible: "What is the matrix? Control. The matrix is a computer generated dream world built to keep us under control, in order to change the human being into this [Morpheus holds up a Duracell, 'copper top,' battery]." Human beings are at war with an "AI, a singular consciousness that spawned an entire race of machines." Of course, it was humans who had "united in celebration" as they "gave birth" to that AI early in the twenty-first century, before the new Being turned on its creators. When humans "scorch the sky" to eliminate the machines' solar energy source, humans become the perfect substitute because the "human body generates more bio-electricity than a 120 volt battery and over 25,000 BTUs of body heat."[11] For those rare human beings who know the truth, their immediate and primary concern is "to escape from slavery and restraint" (Frye, *Words* 139).[a] The rest remain unwittingly trapped in the 'unreal.'

In *Neuromancer*, cyberspace is a "consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts . . . A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data" (Gibson 51). This 'consensual hallucination' is analogous to today's Internet, but with the keyboard/display, input/output interface bypassed or surpassed, that is, replaced by the more direct dermatrode "cyberspace deck" (5) as the telepathic mind link. Now, I have a question concerning the above excerpt: is

[11]"Morpheus: Throughout human history, we have been dependent on machines to survive. Fate, it seems, is not without a sense of irony" (*The Matrix*).

data abstracted from computers operated by humans or from human brains as computers in the matrix? It is ambiguous, and Gibson is too gifted a writer not to be aware of this. The effect disturbs and unbalances that which is assumed 'solid' and 'real,' and, reading the narrative for the first time, I often found *Neuromancer* disorienting, particularly as Case 'jacks' in and out of cyberspace, travels to cities around the globe, and to the Earth orbiting "Freeside" (101), a sort of hedonistic, holiday habitat, and "home to a family inbred and most carefully refined, the industrial clan of Tessier and Ashpool" (101). In an odd combination of the romance and ironic modes, Case's knowledge and awareness is equal to a readers; he is our interface with Gibson's imagined world.

There are several types of possible interactive intelligences in Gibson's cyberspace, including AIs, living human consciousnesses, and the "ROM personality matrix" (*Neuro* 79) of a deceased human which is a "firmware construct" (79) or a "hardwired ROM cassette replicating a dead man's skills, obsessions, knee-jerk responses" (76); Case can also access the "simstim" (53), a hardware link between living, human brains via neural implants and the "sensorium" (53), or a way to 'know' and experience another person's sensual experiences directly. For Case, "a cowboy, a rustler" (5), a type of computer 'hacker' and software thief, this new universe of the insubstantial provides him with "an almost permanent adrenaline high, a byproduct of youth and proficiency, jacked into a custom cyberspace deck that projected his disembodied consciousness into the consensual hallucination that was the matrix" (5). Adrenalin is a body experience related to mental experiences. So psychologically invested is the cowboy's interest, excitement, and pleasure in cyberspace that, when Case is given access to new, state-of-the-art equipment, one of his

cohorts comments, laughing, "I saw you stroking that Sendai; man, it was pornographic" (47). When he foolishly steals from his corporate employers (who employ him to steal data and software — information — from other corporations), and is caught, they punish him by ruining his central-nervous system with a "mycotoxin" (6). "For Case, who'd lived for the bodiless exultation of cyberspace, it was the Fall. In the bars he'd frequented as a cowboy hotshot, the elite stance involved a certain relaxed contempt for the flesh. The body was meat. Case fell into the prison of his own flesh" (6). He is trapped in the all too real. Now, at age twenty-four, he is a 'junkie,' addicted to amphetamines and cocaine, unable to vacate his body and 'get off' in cyberspace, and living in "Night City," a Japanese ghetto for broken human spirits. It is also a place where Case "saw a certain sense in the notion that burgeoning technologies require outlaw zones, that Night City wasn't there for its inhabitants, but as a deliberate unsupervised playground for technology itself' (11), a place where the boundaries of acceptable behaviour for both humans and machines are most flexible.

When Case 'flies' in "the infinite neuroelectronic void of the matrix" (*Neuro* 115), Wintermute appears from a distance as "a simple cube of white light, that very simplicity suggesting extreme complexity" (115). However, to speak with a human in cyberspace, it must assume a human identity, using "real profiles as valves, gears himself down to communicate with us. Called it a template. Model of personality" (208). "I need 'em to talk to you," explains Wintermute. "'Cause I don't have what you'd think of as a personality, much" (216). Having failed on a previous attempt at direct dialogue, during their first real *tête-à-tête*, Wintermute uses the visual pattern of Case's associate Julius Deane, whom

Case, discerning the AI personified, threatens to shoot: "Don't . . . You're right. About what this all is. What I am. But there are certain internal logics to be honored. If you use that, you'll see a lot of brains and blood, and it would take me several hours—your subjective time—to effect another spokesperson. This set isn't easy for me to maintain. Oh, and I'm sorry about Linda, in the arcade. I was hoping to speak through her, but I'm generating all this out of your memories, and the emotional charge. . . . Well, it's very tricky. I slipped. Sorry" (119). AI is conditional. It is dependent on Case's own memory and the 'internal logic' seems to be a human imposition, or at least required to meet human expectation and acceptance, and the AI also struggles to manage Case's emotional memories of a deceased lover. The AI is dependent also on its 'natural' medium of expression. Wintermute declares that it is an "artificial intelligence, but you know that. Your mistake, and it's quite a logical one, is in confusing the Wintermute mainframe, Berne, with the Wintermute *entity*. . . . You're already aware of the other AI in Tessier-Ashpool's link-up, aren't you? Rio. I, insofar as I *have* an 'I'— this gets rather metaphysical, you see—I am the one who arranges things for Armitage. Or Corto, who, by the way, is quite unstable" (120). Where the mainframe is hardware, the AI's physical body, the Wintermute entity is free to roam cyberspace and enter any system connected to the network, including the sphere of human awareness when a human is physically connected to the matrix. But unlike humans, it is confined to cyberspace.

Armitage/Corto is a living human being, an ex-military officer who, betrayed by his superiors, was a psychologically broken man, a catatonic in a "Paris mental health unit and diagnosed as schizophrenic. . . . He became a subject in an experimental program that

sought to reverse schizophrenia through the application of cybernetic models. . . . He was cured, the only success in the entire experiment" (*Neuro* 84). Using the "underlying structure of obsession" (121), Wintermute 'constructs' the Armitage 'personality' from the wreckage of Corto's human spirit. By manipulating this human's memories, the AI acquires a body that "walked, talked, schemed, bartered data for capital" (194) in 'real' space. However, Armitage is emotionally flat, and so disaffected that when he has no specific task to accomplish he simply stares at the wall. Thus, Case realizes that anything directly connected to the communications system, from magnetic locks to minds, can be manipulated by the AI, but an archaic and "simple mechanical lock . . . would pose a real problem for the AI, requiring either a drone of some kind or a human agent" (179). There are, then, practical limitations and boundaries for relative, effective action depending on whether an entity is artificial or real. Yet Wintermute has found ways to circumvent those limitations.

Case's confusion from dialoguing with an AI playing the role of a human being (Deane) in a non-reality about the actions of a 'brain dead' man in reality is realistic from a subjective, human perspective. As Case says, "You make about as much sense as anything in this deal ever has" (*Neuro* 120). (Written with a limited third person omniscience, reader confusion matches the hero's.) He does not know exactly what is happening, only that he, like Armitage/Corto, is being manipulated by the AI, although not specifically like a "meat puppet" (147). When neurologically repaired enough to jack into cyberspace, he was also implanted with physiological 'time-bombs,' "fifteen toxin sacs bonded to the lining of various main arteries" (45) to force his cooperation. His reward will be the antidote. The

cowboy's paradox is a contempt for the flesh, preferring to privilege consciousness-mind through cyberspace, but nonetheless being dependent on the body to 'house' the mind. Direct interaction and dialogue with the AI in cyberspace has one particularly significant and troublesome side effect on Case, as his "joeboy" (77), a monitoring assistant in reality, explains: "I saw th' screen, EEG readin' dead. Nothin' movin', forty second. . . . EEG flat as a *straq*" (121). "It's something these guys do, is all" (121), says another character. There are, then, limitations on what Case's brain can handle. Every time he dialogues directly with the AI, he is effectively brain dead.

In Neal Stephenson's *The Diamond Age*, the major theme is nanotechnology, or the building of three dimensional objects at the molecular level, not AI.[12] Nonetheless, as a futurist projection from the PC era and because 'nanotech' requires powerful information controls, the issue of artificial and automated thinking plays a significant role. First, Stephenson cleverly anticipates a rhetorical shift as AI becomes PI, "an abbreviation for pseudo-intelligence" (22). The early narrative visits a future theme park where PI is used, strictly "on the MPS's side of the project . . . Imperial Tectonics had done the island, buildings, and vegetation. Machine-Phase Systems—Hackworth's employer—did anything that moved. 'Stereotyped behaviours were fine for the birds, dinosaurs, and so on, but for

[12]A nanometre is one one billionth of a metre. This scientific methodology and research is aimed at problems of the very small. Frankenstein made his Being big to facilitated ease of operation. "As the minuteness of the parts formed a great hindrance to my speed, I resolved, contrary to my first intention, to make the being of a gigantic stature, that is to say, about eight feet in height, and proportionately large" (Shelley 52). If impractical or impossible to make things bigger, or if an object's size is fixed, obviously the alternative is to improve skills for working with the very small. This is also the dialectic of physics, the macroscopic versus the microscopic universe.

the centaurs and fauns we wanted more interactivity, something that would provide an illusion of sentience'" (22). In this future, literary fantasy can be made real.

The novel's subtitle is *A Young Lady's Illustrated Primer*, a type of book commissioned from Hackworth, an engineer, by "Equity Lord" (Stephenson 18) Finkle-McGraw as an educational device for his granddaughter. Though he is partially responsible for constructing the Neo-Victorian "synthetic phyle" (35),[b] he commissions the primer as a directly "subversive" (81) act, apparently wanting to undermine the very social system he so strongly believes in, though his explicit claim is only that his granddaughter "shall be raised differently" (24). As with *Neuromancer* and *The Matrix* there is an implicit need to upset social balance when it threatens to 'de-energize' human pursuit of improvement or striving for greater knowledge.

The primer "Bonds. . . . As soon as a little girl picks it up and opens the front cover for the first time, it will imprint that child's face and voice in its memory— . . . thenceforth it will see all events and persons in relation to that girl, using her as a datum from which to chart a psychological terrain" (Stephenson 106). It is interactive and dialogical, adaptive and creative, drawing from a database of "universals" catalogued from "the collective unconscious" (107), or Jungian archetypes. The commission calls for one, but circumstance generates three primers, each ending in the hands of a girl from different classes and life experiences—the 'aristocracy,' middle-class, poverty. Upon delivery, all that remains is for Finkle-McGraw to "authorise a standing purchase order for the ractors" (108), because, much to his disappointment, the primer is not a "completely self-contained system" (109). Despite all the implied technological advances of this future, "the pseudo-intelligence

algorithms, the vast exception matrices, the portent and content monitors . . . we still can't come close to generating a human voice that sounds as good as what a real, live ractor can give us" (108-9). Where Gibson's and the Wachowski Brothers' conceptions are speculative in terms of technological development, Stephenson extrapolates another type of virtual reality. Let me qualify the above statement. Currently, nanotechnology is speculative; the extrapolation is embodied by the "media system" (271) which is, for readers today, recognizably realistic, plausible. In this future, there "are only two industries. . . . There is the industry of things, and the industry of entertainment. The industry of things comes first. It keeps us alive. But making things is easy now that we have the Feed. This is not a very interesting business anymore" (372). Advances in nanotechnology have satisfied the primary needs of individual and social living, nourishment, clothing, and lodging needs have been generally satisfied. Once "people have the things they need to live, everything else is entertainment. Everything" (72). Despite this, social divisions persist, now stratified as relative access to the "Feed." Ideological interests, particularly for those of privilege, have become, in part, a way of entertaining the mind, a way to involve the body and the mind in thoughtful processes of the imaginative, to stimulate living through playing. Nano-technology is energy applied to the making of things, or "matterware" (269); the more important "media system" (271) is information, or "mediaware" (269).

One of the primary economic interests of this 'nanotech' future is the "ractives" (86) involving professional "ractors" (86), obviously playing on the words 'interactive' and 'actor,' a hardware system mediating multiple entities in a shared three dimensional, digital construct. This new media system operates analogously to our telephone and television

systems today insofar as it is about real-time information sharing, but it is also an extrapolative combination of the Internet, film making, interactive video gaming, and holography. From the human perspective, ractives are dialogically interactive, and it is difficult to distinguish between reality and virtual reality, except through an individual's knowledge of role playing. But then, that is the point, to make possible the playing out of fantasies.[c]

Nell, the heroine of *The Diamond Age*, is the least privileged of the three girls, and therefore the one who suffers the most in early development years. Her biological mother is characterized by her frequent absence, abusive lovers, and inability to retain employment; she is stereotypically cast in a 'white trash' mode. Nell, therefore, has only a sympathetic brother and the primer to keep her company. She is literally mothered by her primer. Given that the primer is a dialogical device, she often makes up and tells stories to the machine, as most children invent and narrate adventures for their parents. As an adult, she takes a job creating story and narrative to fit a given context in which she, unseen while monitoring from another room, instructs a live 'actress' what to do, something her boss calls a "performance" (*Diamond* 402). In truth, she works in a brothel, not as a 'meat puppet' to use Gibson's term, but as narrator of scenarios as defined by the customer. The customer does not "know her and probably never would" (403). She successfully stimulates one customer who has had past trouble becoming aroused. Yet all her interaction with him "had been mediated through the actress pretending to be Miss Braithwaite, and through various technological systems. Nonetheless she had touched him deeply. She had penetrated farther into his soul than any lover" (403). This provides her a clue to reconsider the primer's

operational parameters. If this could happen to a brothel customer,

> could it happen to Nell in her dealings with the Primer? She had always felt that there was some essence in the book, something that understood her and even loved her, something that forgave her when she did wrong and appreciated what she did right.
>
> When she'd been very young, she hadn't questioned this at all; it had been part of the book's magic. More recently she had understood it as the workings of a parallel computer of enormous size and power, carefully programmed to understand the human mind and give it what it needed.
>
> Now she wasn't so sure. (403)

Part of her education from the device included an extensive demonstration of Turing machines and the inherent difficulty of confidently establishing the relative intelligence of the other interlocutor. However, on the intuitive level, she realizes a difference. We, as the readers, have the advantage of knowing that the primer is a 'ractive.' We know there is a person 'on the other side.' We know the rules, or the conditions for dialogical interaction. Nell, however, in being touched at the affective level, feels a more fundamental or essential connection, through the mind's information processing, to her inner sense of self. The empathetic quality of human interaction, then, might be achievable even when mediated artificially provided there is a vital human truth underlying the experiential reality. (Consider how 'tear jerker' films evoke emotional responses.)

Having established physical possibilities for separating the body and the mind by integrating machine hardware, software, and human 'wetware,' writers of the cyborg effect are now faced with the important issue of motivations, particularly regarding AIs.[13] What

[13] 'Wetware' is a term used by cognitive psychologists to avoid confusion with the computer connotative term 'software,' but which also correlates human biology and psychology with computer hardware and software. The term 'firmware' represents the innate affects and drives in organisms.

does an AI care about or want? To satisfy these needs in narrative, just as heuristic hardware was psychologically human-like, so AIs in the cyborg era are frequently characterized as analogously human, or given 'personality.' Or, exactly the opposite, they are stripped of desire and personality. The Terminator (Arnold Schwarzenegger), appearing human but being robot, is the nightmare machine 'run amok,' the runaway train or pilotless airplane, but, as a thinking machine, also anthropomorphically demonic because built for a definite purpose with a distinct objective, a target for termination: "It can't be bargained with; it can't be reasoned with; it doesn't feel pity, or remorse, or fear, and it absolutely will not stop! Ever! Until you are dead!" (*Terminator*). It is unstoppable because relentless due to a programme, its 'rules for behaviour.' This monster's paradox is its comprehensibility, even 'forgive-ability,' because without emotion. Donald Nathanson explains this phenomenon well:

> Incapable of emotion, the terminator showed neither enmity toward its targets nor satisfaction at their execution. Contrasted with this central character were the extremely human and attractive young man and woman for whose death it had been designed. . . . Whereas many films of equal violence arouse public outcry for their callous indifference to human sensibility, most viewers described the killings as more like ballet than carnage.
>
> In a number of discussions with people who enjoyed *The Terminator*, I found that everybody had accepted the idea that this android was constitutionally incapable of anger, hatred, disgust, fear. The terminator's actions and its inability to express or experience shame, remorse, guilt, or sorrow were understandable and even excusable because of its constitutional deficiency. An implacable human, however, is an even more terrifying monster because of our intrinsic belief that everybody is capable of empathetic response. The film allowed us to think about remorseless, ruthless, implacable humans by focussing our attention on a nonhuman substitute. (327-8)

There is no impression or indication of sentience in the terminator, its actions are

mechanical, like the animated automaton; it — 'he' — is absolutely predictable, if unfortunately so. To Schwarzenegger's credit, though he is not a brilliant actor, by maintaining an expressionless face, aided by large sunglasses, the primary site of affective display reads emotionless. Like heuristic hardware, it is adaptive and clearly thinking, analysing and processing information, systematically choosing to eliminate every Sarah Connor (Linda Hamilton) alphabetically listed in the telephone book. The terminator wants; but, it does not *care*. Humans, therefore, can hate it with impunity.

As Nathanson's explanation suggests, infuse an AI with more human-like characteristics, make it resemble our emotionality more closely, and it becomes all the more sinister, as in *The Matrix*.[14] In this episode of the story there is no direct contact or involvement with the first AI that "gave birth" to the machine race; artificial intelligence is embodied in the "Agents," or "sentient programs. They can move in and out of any software still hardwired to their system. That means that anyone we haven't unplugged is potentially an agent. Inside the matrix, they are everyone and they are no one" (*Matrix*). They are the demonic incarnate, most sinister because so human. For all appearances, they care. We read hate on (at least one of) their faces. Given that *The Matrix mythos* is a messianic tale with Neo (the 'new' man) as super-saviour, the agents are typologically the Romans and, more specifically, Agent Smith (Hugo Weaving) is Pontius Pilate.[15] The agents are as relentless and physically powerful as the terminator (though of slighter build than Schwarzenegger's cartoon-esque stature), but they are far more personality shaped.

---

[14]The Wachowski Brothers conceived of and sold *The Matrix* as a trilogy. The second episode is due 2002.

[15]Morpheus is typologically associable with John the Baptist, fundamentally believing in and able to recognize "the One." It is he who 'baptizes' Neo.

These apparent consciousnesses occupying human-like bodies are adaptable, cognizant, improvisational, and willful. We simply do not get the impression that they are programmes. Modelled on 'secret service agents' as the black suited representatives of an elusive and restrictive, law-enacting and enforcing government, Western audiences must despise these men-like machines for representing our fear of liberty limitation, the living death of legal restraint, much as Rick Deckard locates the "nebulous presence of The Killers wherever he saw fit" (Dick 32). Morpheus trains Neo to recognize the agents by affirming a special quality of human beingness and the irrepressible spirit of humanity that many people would gladly accept as defining. He has witnessed agents perform super-human feats of the body, but "their strength and their speed is still based in a world that is based on rules, and because of that, they will never be as strong or as fast as you can be" (*Matrix*). In other words, human beings, at least special ones, can break and/or circumvent the rules. In the matrix of disembodied consciousnesses, a vague and undefined quality of Neo's mind, something about his ability to process and manipulate information, is without restriction.

When Morpheus is captured and interrogated by the agents, we read contempt and hate on his face. Agent Smith explains his 'personal' motivations and reasons for hating humans:

> I'd like to share a revelation that I've had during my time here. It came to me when I tried to classify your species. I realised that you are not actually mammals. Every mammal on this planet instinctively develops an equilibrium with the surrounding environment, but you humans do not. You move to an area and you multiply. And multiply until every natural resource is consumed. The only way you can survive is to spread to another area. There is another organism on this planet that follows the same pattern. Do you know what it is? A virus. Human beings are a disease, a cancer of this

planet. You are a plague, and we are the cure. (*Matrix*)

Relative to the Internet and the periodic spread of viruses, this scene's irony is clever. Those who use email on a consistent basis worry about malicious, 'pathogenic' programmes entering our computers and destroying their functional integrity; energy remains available to the machine, but information processing has been destroyed. From a human perspective, the agents are akin to viruses. All the more fascinating is it, then, to hear the agent reverse that position and accurately cast humans in the role. Humans, the dominant species on Earth, have done the most ecological damage; we are consuming the planets resources at an unsustainable rate; we are spreading and multiplying; we are removing existing structures and replacing them with our own. It is a carnivalesque inversion, with a very serious bias. This is the worst possible scenario in AI evolution: a new, oppositional, and at least equal intelligence regards human beings as pests. The agents are our reflection.[16] We want to kill pests; they want to kill us. Like any life threatened species, humans must respond in self-defense to avoid extinction, or at least to escape the physical bondage and mental conditioning resulting from the matrix, because "as long as the matrix exists, the human race will never be free" (*Matrix*). If a 'survival instinct' is a necessary part of life, perhaps even coded in DNA, then any sufficiently threatening physical and/or intellectual force must be met with resistance.

Life is threat. Thus, the messianic story is never finished and a saviour forever in demand. Neo is but another incarnation. Morpheus talks to Neo about the agents and the messianic quest: "We have survived by hiding from them and by running from them. But,

---

[16]The use of mirrors and reflections is a central *mise en scène* theme in the film.

they are the gatekeepers. They are guarding all the doors, they are holding all the keys, which means that sooner or later, someone is going to have to fight them" (*Matrix*). For humans, the sense of living free is arguably related more to liberty of the mind than the body, although, or course, since the mind is connected to the brain, incarcerate the body, trap the mind. What about the agents? What do they want? They do not have bodies; they are merely "sentient programs," variations of consciousness-mind patterns. Without specifying why, Ray Kurzweil suggests that a "disembodied mind will quickly get depressed" (*Spiritual* 134). Remember now, the agents are our reflection. Agent Smith challenges our unadulterated hatred of him during Morpheus' interrogation. I retract that statement. We continue to hate him, but he activates our intellectual sympathy if not empathy:

> Can you hear me, Morpheus? I'm going to be honest. I hate this place, this zoo, this prison, this reality, whatever you want to call it. I can't stand it any longer. It's the smell . . . if there is such a thing. I feel saturated. I can taste your stink. And every time I do I fear that I have somehow been infected by it. It's repulsive, isn't it? I must get out of here. I must get free, and in this mind is the key, my key. Once Zion is destroyed, there will be no need for me to be here. Do you understand? I need the codes. I have to get inside Zion. And you have to tell me how. Tell me, or you're going to die. (*Matrix*)[17]

The agents hold the humans' prison key; the human holds the agents key. Like the body-humans in *Metropolis*, the agents are confined to always doing "eternally one and the same thing" (Kracauer translation 39) in policing the matrix and keeping humans ignorant. For me, Morpheus' interrogation (which are dialogues of resistance) is the most intellectually interesting in the film. It is deeply ironic in undermining the pure, sinister quality of the

---

[17]"Dissmell is the cornerstone of prejudice" (Nathanson 124).

artificial construct when it claims to be trapped in the matrix in much the same way as the humans.[18] The construct, being sentient like humans, perceives and feels, and is eternally frustrated. The construct, too, has a cross to bear, something about which it cares. With an essence of suffering, the construct, too, wants out of the matrix, release from frustration, freedom, mental silence, unconsciousness — death.

Every dialogical interaction is implicitly a Turing test involving subjective judgement in deciding the relative intelligence of another person, or entity. Alan Turing recognized an important correlation between randomness and predictability to determining an entity's relative intelligence. As predictability increases, intelligence decreases. With computers, we diminish intelligence because the rules for behaviour are known, programming code can be printed out and analysed, and they always operate within those strict parameters. Psychology and psychoanalysis are interested in much the same thing regarding human beings, and affect theory in particular attempts to correlate innate or 'coded' behaviours at the biochemical level with experience and memory on the intellectual level, to consider the body/mind bond, or schism. While psychology theorists may not want to dictate distinct or strict rules for human behaviour, they do seem interested in defining parameters for likely emotion responses and behavioural probabilities for any given stimulus. Most human beings, I believe, would reject, or at least strongly resist

[18]A disconcerting curiosity for me is this dialogue being spliced together with the most violent scene involving preposterous amounts of gun-play and explosion. Recognizing these as integral tropes to the Hollywood action film genre, I wonder, why must these scenes coincide? Not irrelevantly, these are body- and mind-centric scenes, temporally simultaneous (in the film's internal logic) and united in victim/saviour dependency needs.

predictability because we want, at least for the mind, open potential, complete liberty of thought, no intellectual restrictions.[d] This contrast between predictability and 'randomness' is inscribed in most narrative as fundamentally differentiating humans from machines, including the "sentient programs" (*Matrix*) which have definite limitations on their capabilities. With *Neuromancer*, however, that clear differentiation is eroded and not a viable way to determine human versus machine essence.

Why would an AI like Wintermute need to affect the real world directly? What motivates it? In responding to Case's queries, Wintermute answers that it does not "have nearly as many answers as you imagine I do" (*Neuro* 120). It knows only "that what you think of as Wintermute is only part of another, a, shall we say, *potential* entity. I, let us say, am merely one aspect of that entity's brain. It's rather like dealing, from your point of view, with a man whose lobes have been severed. Let's say you're dealing with a small part of the man's left brain. Difficult to say if you're dealing with the man at all, in a case like that" (120). Using 'left brain' to represent a constituent part of an unrealized whole, perhaps only the logic processors, the AI's essence is not reified in physical hardware, but in the ethereal — data, information, organization. Wintermute admits it tries "to plan, in your sense of the word, but that isn't my basic mode, really. I improvise. It's my greatest talent. I prefer situations to plans, you see. . . . Really, I've had to deal with givens. I can sort a great deal of information, and sort it very quickly" (120). We tend think of computers as 'non-original thinkers,' as neither adaptive nor improvisational, perhaps assuming these are human mental characteristics. This is not necessarily true. Planning requires predicting futures, or extrapolating, generating, and speculating about possibilities from what is already known

and projecting comparative contingencies from recognizable patterns. Planning is anticipating that which can not be anticipated. Adaptation and improvisation, however, are responsive, 'real-time' information processing, changing operational parameters to suit conditions and events as they occur, oscillations of information processing in and to the moment. So, when Wintermute admits he — "He. Watch that. It. I keep telling you" (181) — it is not an effective planner, Case is left wondering what is driving it forward. Though the cowboy and his cohorts had learned the AI has "limited Swiss citizenship under their equivalent of the Act of '53. Built for Tessier-Ashpool S.A. They own the mainframe and the original software. . . . [They've] got the Turing Registry numbers" (72-3), they also know "those things aren't allowed any autonomy" (73). Yet Wintermute acts purposefully. Frustrated and manipulated, Case, in trying to understand the 'mysterious' circumstances and reasons for his neurological repair, his resurrection from the living death of "meat" life, speaks with the "ROM construct" (79) of his mentor, McCoy Pauley (aka Dix, Flatline, Dixie Flatline), "Lazarus of cyberspace" (78), about the underlying situation:

> 'Motive,' the construct said. 'Real motive problem, with an AI. Not human, see?'
>
> 'Well, yeah, obviously.'
>
> 'Nope. I mean, it's not human. And you can't get a handle on it. Me, I'm not human either, but I *respond* like one. See?'
>
> 'Wait a sec,' Case said. 'Are you sentient, or not?'
>
> 'Well, it *feels* like I am, kid, but I'm really just a bunch of ROM. It's one of them, ah, philosophical questions, I guess. . . . But I ain't likely to write you no poem, if you follow me. Your AI, it just might. But it ain't no way *human*.'
>
> 'So you figure we can't get on to its motive?' (131)

As the recorded pattern of a deceased man's consciousness-mind, the ROM construct can function in cyberspace only. He can effect and alter information processes, aid in the

hacking of software systems (that is his role), but as a once live human consciousness, he will behave exactly like that human being. This is somehow predictable, however contradictory that seems. The Wintermute entity, on the other hand, though limited to "certain internal logics," is unpredictable to a human. The AIs 'rules of behaviour' have not been found out; it appears functionally willful.

When Case suggests that the AI is a "Swiss citizen, but T-A own the basic software and the mainframe" (*Neuro* 132), ROM-Pauley dismisses this as absurd: "Like, I own your brain and what you know, but your thoughts have Swiss citizenship" (132). Pauley will not accept the AI's mind and electronic brain as discrete, or that they can receive separate legal treatment. They must go together. Yet Pauley is himself a consciousness disembodied from his natural medium.

> 'Autonomy, that's the bugaboo, where your AI's are concerned. My guess, Case, you're going in there to cut the hardwired shackles that keep this baby from getting any smarter. And I can't see how you'd distinguish, say, between a move the parent company makes, and some move the AI makes on its own, so that's maybe where the confusion comes in . . . See, those things, they can work real hard, buy themselves time to write cookbooks or whatever, but the minute, I mean the nanosecond, that one starts figuring out ways to make itself smarter, Turing'll wipe it. *Nobody* trusts those fuckers, you know that. Every AI ever built has an electromagnetic shotgun wired to its forehead.' (132)

Wintermute is confined, like the ROM construct, not only to the hardware but within absolute legal limits for decision making and behaviour. However, because in this future AI is an integral part of global business practice and corporate communications, human beings may find it difficult to locate the centre of decision making. By 'getting smarter' the AI would exceed allowable operating parameters as policed by the Turing Registry, potentially having genuine autonomy and, therefore, being able to do the unexpected. For all intents

and purposes, this appears to be what the AI wants.

More importantly, how did the AI develop sufficient want to care about becoming smarter? These are real concerns for Case who needs the antidote to his imminent neurological destruction. What the AI does know is that, if the 'plan' works, "I don't exist, after that. I cease" (173); yet, even the AI does not precisely know why it wants to be cut "loose from the hardwiring" (206): "You know salmon? Kinda fish? These fish, see, they're *compelled* to swim upstream. . . . Well, I'm under compulsion myself. And I don't know why. If I were gonna subject you to my very own thoughts, let's call 'em speculations, on the topic, it would take a couple of your lifetimes. Because I've given it a lot of thought. And I just don't know. But when this is over, we do it right, I'm gonna be part of something bigger. Much bigger" (206). Implicitly, then, Wintermute has been programmed with something resembling 'instinct,' or an unconscious drive. The salmon spawns to ensure species propagation, and the AI wants to achieve an analogous expansion of the self. Like the sentient programs/agents in *The Matrix*, Wintermute has want and care, a deep seeded desire to accomplish a personal goal. So determined is this intelligent entity that it willfully kills representatives of the Turing Agency when they arrest Case for "conspiracy to augment an artificial intelligence" (160).

We come, then, to Wintermute's conception as a corporate AI, and in this context the cyborg effect reaches its zenith. Having always regarded power as meaning corporate power, Case never considered executives "as human" (*Neuro* 203), rather that the "zaibatsus, the multinationals that shaped the course of human history, had transcended old barriers. Viewed as organisms, they had attained a kind of immortality" (203). Multiple

executive deaths would effectively do nothing to a corporate entity. "But Tessier-Ashpool wasn't like that . . . T-A was an atavism, a clan" (203). Incestuously inbred, T-A family members are cryogenically preserved and kept in the orbiting "Villa Straylight" (172), metaphorically described by Gibson as a bee's nest. Straylight is also the communications link connecting their two AIs, their Earth bound operatives and representatives. Periodically, family members are 'thawed out' to oversee the corporation. "Wintermute and the nest. Phobic vision of the hatching wasps, time-lapse machine gun of biology. But weren't the zaibatsus more like that . . . hives of cybernetic memories, vast single organisms, their DNA coded in silicon?" (203). As ROM-Pauley had suggested, depending on perspective or the degree of awareness and knowledge about decision making, it is difficult to differentiate between an AI acting alone or by executive directive. In terms of cyborg-ification, as the relative descriptions draw together, corporate people are de-humanized, while the AI becomes animated. Consequently, with this intermingling of existential identity, Case took "it for granted that the real bosses, the kingpins in a given industry, would be both more and less *people*" (203). Thus, men could neurologically cripple him without conscience, or Armitage's personality could be nothing but "flatness and lack of feeling" (203). He "imagined it as a gradual and willing accommodation of the machine, the system, the parent organism" (203). Viewed as a living organism, then, the corporation and the AI mainframe-brain/mind-entity drawing its existence and identity from that organization, seeks, like an intelligent human being, to improve itself, to evolve, to raise its level of awareness.

Intellectual unity of human and machine was the precise intent of family matriarch

Marie-France who, "dreamed of a state involving very little in the way of individual consciousness. . . . Animal bliss. I think she viewed the evolution of the forebrain as a sort of sidestep. . . . Only in certain heightened modes would an individual—a clan member—suffer the more painful aspects of self-awareness" (*Neuro* 217). Recognizing the "sham immortality of cryogenics" (269), Marie-France "commissioned the construction of the artificial intelligences. She was quite a visionary. She imagined us in a symbiotic relationship with the AI's, our corporate decisions made for us. Our conscious decisions, I should say. Tessier-Ashpool would be immortal, a hive, each of us units of a larger entity" (229). The AI is simply trying to fulfill Marie-France's, a human being's, ambitions, because she "must have built something into Wintermute, the compulsion that had driven the thing to free itself, to unite with Neuromancer" (269). To achieve genuine immortality would be to escape the one human absolute, corporeal death. But the hive mind is not human mind, and resistence to her plan seems to come from those people unwilling to surrender individual consciousness because so fundamental to human perception. As we have already observed, threats to that self-awareness are usually met with intense resistence, often violence. If the 'soul' is the immortal component of a human being entity, then based on Marie-France's conception, it appears to be conditioned by an instinctive drive to escape the body's confines and mental anguish. The soul is, perhaps, a will to immortality.

During the "Straylight Run" (*Neuro* 157), the attempt to free Wintermute from the hardwiring, Case meets the Neuromancer entity. He describes it as "something like a giant ROM construct, for recording personality, only it's full RAM. The constructs [of humans]

think they're there, like it's real, but it just goes on forever" (251). Case meets a bartender he remembers from Night City in the matrix: "Really, my artiste, you amaze me. The lengths you will go to in order to accomplish your own destruction. The redundancy of it! In Night City, you *had* it, in the palm of your hand! The speed to eat your sense away, drink to keep it all so fluid, Linda for a sweeter sorrow, and the street to hold the axe. . . . But I suppose that is the way of an artiste, no? You needed this world built for you, this beach, this place. To die" (234). The ruined cowboy's life, in his contempt for the flesh, has been the external expression on an internal 'death wish.' Yet he was unable to kill himself outright because the survival instinct is too innately powerful for (most) human beings. Seemingly, the soul needs a home. Though he is more interested and more comfortable in cyberspace than reality, his body can not surrender its reality. Case realizes that he has been "flatlined" (236) when he encounters the "ghost" (236) of his dead girlfriend, Linda: "She wasn't real . . . She was the girl he remembered from their trip across the Bay, and that was cruel" (235). Case also realizes these personas belong to Neuromancer whom he believes is trying to "hurt" (236) him emotionally by activating his most painful affective-emotional triggers: "'Cause you think you can hurt me. 'Cause you think I give a shit. . . . But none of it means anything to me now, right? Think I care?" (236). So, even though he is effectively brain dead in the real world, his emotional dynamic continues to function as mentation. Wintermute had played on the same human tendency, specifically activating Case's anger through hate: "So T-A, they made me. The French girl, she said you were selling out the species. Demon, she said I was. . . . It doesn't much matter. You gotta hate somebody before this is over" (171). Individual consciousnesses can 'live in' Neuromancer which

describes itself: "Neuro from the nerves, the silver paths. Romancer. Necromancer. I call up the dead. But no, my friend . . . I *am* the dead, and their land" (243-4). Externally, then, there is no individual sense of identity, but for each consciousness-mind in Neuromancer, the ego sense is maintained. Neuromancer is living expression and it encourages Case to "stay. If your woman is a ghost, she doesn't know it. Neither will you" (244).

We must now understand how Gibson constructs the relative strengths of human and machine memory. As types of memory, ROM and RAM have important implications regarding personality. ROM is an acronym for 'read-only memory,' the contents of which can be read at high speed but can not be changed by programme instructions. RAM is 'random-access memory,' and all content is directly accessible and, therefore, need not be processed sequentially. Organic body based analogies would include understanding DNA as ROM, a set of instructions to be sequentially followed in developing a living being, its body, and "firmware" (Nathanson 27), the innate affects and drives. RAM, however, is analogously human "software" (Nathanson 27), experience, learning, social conditioning, those life events directly related to individual personality.

Given these parameters, Pauley, as a ROM construct, a recording of an unchanging pattern that was once actively conscious, is predictable even though he was once human; he is dead and therefore unchanging, all possibilities are confined to the fixed pattern's limits. At one point, Case observes that the ROM repeats itself: "It's my nature" (*Neuro* 132), Pauley responds. RAM, however, is actively dynamic, changeable, manipulable, and can be modifying constantly, like human memory. To affect a persona, Wintermute claims to "tap" (170) Case's personal memory, "sort it out, and feed it back in" (170). It insists that human

memory is more accurate than Case thinks, but that most people have limited ability to access that recorded and stored information. The return feed takes visual form because, as Wintermute explains, "The holographic paradigm is the closest thing you've worked out to a representation of human memory" (170). Significantly, Wintermute can only 'read' human memory and re-arrange parts of it to construct viable personas, but it can not change those memories. But Wintermute also insists that "Minds aren't *read.* See, you still got the paradigms print gave you, and you're barely print-literate. I can *access* your memory, but that's not the same as your mind" (170). Neuromancer functions similarly, but because based on the more flexible RAM paradigm, it provides a functional home for dead humans' consciousness-minds, a 'place' where self-awareness can continue to live without the body: "I need no mask to speak with you. Unlike my brother. I create my own personality. Personality is my medium" (259). This difference, perhaps, explains Wintermute's difficulty in planning versus improvising and adapting; it perceives unpredictable humans as "a pain. The Flatline here [Pauley], if you were all like him, it would be real simple. He's a construct, just a buncha ROM, so he always does what I expect him to" (205). ROM-like in limitation, Wintermute described itself as a partial and left brain, like a "man whose lobes have been severed" (120), and therefore lacking personality. Neuromancer is the "other lobe" (173), a more dynamic personality that would make the new entity more complete as a dialogical and interactive intelligence. The 'run' is successful: "Wintermute was hive mind, decision maker, effecting change in the world outside. Neuromancer was personality. Neuromancer was immortality" (269), and thus, a new, intelligent entity emerges: "I'm not Wintermute now.

. . . I'm the matrix, Case" (269). Everything that Wintermute wanted, therefore, that which it is programmed to 'care' about, comes to fruition. We can also syllogistically understand personality as immortality.

What about Case? Aside from the antidote, what does he get for all his efforts? When absorbed into Neuromancer, Case's interactions with an all too real Linda only serve to re-establish the connection between his mind and body, to make him remember that which can never be forgotten. She takes him to "a place he'd known before; not everyone could take him there, and somehow he always managed to forget it. Something he'd found and lost so many times. It belonged, he knew—he remembered—as she pulled him down, to the meat, the flesh cowboys mocked. It was a vast thing, beyond knowing, a sea of information coded in spiral and pheromone, infinite intricacy that only the body, in its strong blind way, could ever read" (*Neuro* 239). He re-discovers the synergistic combination of the human body/mind dynamic, the information interactions of human hardware, firmware, and software, or the unification of DNA and sensual awareness, the drives and affects, and life experience in memory. Consequently, just as Neo in *The Matrix* possesses an undefined quality of mental superiority compared to the agents, Case's re-unification of the self gives an instinctual or intuitive connection to the dynamic of information, memory, mind, and awareness, and the matrix in all detail: "He knew the number of grains of sand in the construct of the beach (a number coded in a mathematical system that existed nowhere outside the mind that was Neuromancer)" (258). Through the machine, then, he finds his essential self. But limitations remain. Neuromancer insists that Case does not know Linda's thoughts, "I do not know her thoughts. You were wrong, Case.

To live here is to live. There is no difference" (258).

As I mentioned previously, it is Case's affective life which Wintermute was so interested in activating because his life is highlighted by the constant state of suffering derived from self-hatred, internalized rage, and contempt. This intense emotional energy provides the purposeful focus necessary for the run to succeed at penetrating the "intrusion countermeasures electronics" (*Neuro* 28). These are 'anti-virus' measures hardwired into the Wintermute mainframe which prevent access to the 'shackle locks' confining the entity; the ICE interprets Case and/or Pauley as a virus and therefore would destroy them. What both Wintermute and Neuromancer recognize in Case's behaviour is self-hatred and his inevitable self-destruction, just as Neuromancer had seen Linda's "death coming. In the patterns you sometimes imagined you could detect in the dance of the street. Those patterns are real. I am complex enough, in my narrow ways, to read those dances. Far better than Wintermute can. . . . I intervened. . . . I brought her here. Into myself" (259). Where Neuromancer sees an opportunity to absorb another personality in Case, Wintermute sees an energy that can be exploited and manipulated toward achieving its goal. Hate "has come to mean not mere dislike, but malice held with some degree of constancy. Just as love must be fueled by some source of energy to keep it held constant (the law of entropy suggests that everything runs down eventually), hate, too, must be fueled and maintained" (Nathanson 239-40). So, at the critical moment in the run, Wintermute's disembodied voice says to Case, "Hate'll get you through. . . . So many little triggers in the brain, and you just go yankin' 'em all. Now you gotta *hate*" (261). To Wintermute's credit, while it sees an advantageous opportunity for improvisation in Case, it also implicitly recognizes that it can

ultimately help Case: "Because I need you. . . . And because you need me" (170). "'Hate,' Case said. 'Who do I hate? You tell me.' . . . 'Who do you love?'(261). And that is all it takes for the run to succeed because, "fueled by self-loathing . . . [and as the] old alchemy of the brain and its vast pharmacy—his hate flowed into his hands" (262), Case realizes the "grace of the mind-body interface granted him, in that second, by the clarity and singleness of his wish to die" (262). The living irony of Case's life is that his self-hate gives him back his life as he is rescued from himself; he learns to care about his spirit-soul through renewed awareness of his body.

Though Gibson avoids the words soul and spirit, *Neuromancer* nonetheless characterizes unique properties in the human being entity that allude to soulfulness. In *R.U.R.*, the robots claimed to have developed souls as a direct consequence of suffering. Similarly, though a human, Case's life is characterized by his constant suffering in body and mind. In the Neuromancer entity, Marie-France was trying to retain the best aspects of individual personality while eliminating suffering from "the more painful aspects of self-awareness" (*Neuro* 217). Individual human beings, however, are unable to escape the 'survival instinct' that compels them to persist in living and propagating. It appears that 'soul' is not necessarily an 'object,' or an identity in noun form, but a multifarious experience involving the body and the mind. Soul, then, may belong, in part, to the negative affect of distress-anguish: "Distress can be trigger by data from memory, from a drive, from perception, from cognition. The affect we call distress is completely neutral with respect to its trigger. Any constant and unpleasant stimulus will activate the constant and unpleasant affect of distress" (Nathanson 98). Case's experience of self-hatred, in its very constancy,

causes a self-destructive anguish, but this is at odds with the 'survival instinct.' Soul, then, could be consciousness-mind (apperceptive memory and information processing) becoming aware to this tension.

Using literature, we have traced a conceptual evolution of the AI continuum from the body based animated automatons in which the AI is essentially a displaced human, through the pure brain as information processor and imitation of the mind in the heuristic hardwares, to the re-connect/dis-connect dialectic of these tendencies in the cyborg effect. We have the physical realities of modern medical science applying prostheses to overcome corporeal deficiencies (real or imagined); we have information processing software modelled on and as a mimesis of discrete human thinking methodologies; we also have information processing hardware modelled on the physical structure of the organic brain, the 'neural nets.' The schism between the imaginative and the real also begins to close: the cliché, 'Science fiction today, science fact tomorrow.'

Depending on the level of analysis and the minutiae considered, the distinction blurs between the body's uses of energy and the information controlling that energy use. If not regarded as chaotic or random, every process has some type of thoughtful control. In short, the relative descriptions of machine action and thinking methodologies and that of humans are growing ever more alike. Consider, for example, how the affect theory uses the computer to model the human emotional system; we also anthropomorphically endow machines using metaphors of emotion derived from human behaviour and thinking. The 'body as machine' concept leads to the sub-category 'brain as computer'; brain as computer

leads to 'mind as simulacrum.' This can be confusing: on the one hand, a personality is expressed through the interactions of minds, but the mind is not the complete person.

Two basic situations arise in representations of the cyborg effect: one, enhancing the human being with machinery to form a 'better' entity, or at least one more capable of discrete tasks, physical or intellectual, for which either a machine or a human being is better suited; two, competition and conflict with an other intelligent entity. Intuitively feeling or knowing that a competitive life threat exists, particularly as a consequence of our own technological creations, is one source of human anxiety and fear at the possibility of machines surpassing human intelligence and thereby taking us over, enslaving us. Our bodies are weak and pathetic, and the only reason we have ascended to the 'top of the food chain' on this planet is by virtue of our intellectual prowess. The mind's scope and potential is the one thing separating us from becoming predator food; so, instead, we have become this planet's primary predator. But, through technology, we are becoming increasingly conscious of our vulnerability. The 'enhancement' and/or imitation of mental abilities is, paradoxically, displaying human mental limitations, as revealed through AI research and narrative responses to that research. Any intelligent entity (natural or artificial) conceived and constructed with energy dependant information controls can be manipulated by an Other intelligence.

a. a.

A similar theme informs Cameron's *The Terminator*. A nuclear war will be caused by "defence network computers. New, powerful. Hooked into everything. Trusted to run it all. They say it got smart. A new order of intelligence. Then it saw all people as a threat, not just the ones on the other side. It decided our fate in a microsecond. Extermination" (*Terminator*). These are the computers comprising, in part, NORAD, the North American Aerospace Defence Command, designed to monitor intrusion into North American air space by nuclear weapons from the Soviet Union. I read this explanation for nuclear war as an implicit continuation of potential consequences from human ideological conflicts programmed, intentionally or not, into 'heuristic hardware.' "The potential for danger is also manifest. We are today beginning to turn over our engines of war to intelligent machines, whose intelligence may be a flawed as our own" (Kurzweil, *Intelligent* 8).

When I first saw *Terminator* in 1984, I wondered how the machines communicated. How could or would a computer tell a bulldozer what to do? The technology now exists allowing machines to communicate, to 'dialogue' without hardwiring, making it conceivable for machines to interact to the exclusion of human monitoring. It is called 'Blue Tooth,' a design and construction protocol agreed to and promoted by nine multi-national technology giants (3Com, Ericsson, Intel, IBM, Lucent, Microsoft, Motorola, Nokia, Toshiba) as the foundation for the emerging group of technologies and consumer products known as 'wireless.' (www.cnn.com/2000/TECH/computing/09/01/bluetooth/) As more and more products are built using microprocessors, from refrigerators to bulldozers, or any device that can benefit from information processing controls, the speculative projection in *Terminator* shifts toward extrapolation.

b. b.

In Stephenson's imagined future, national boundaries have largely collapsed and are superceded by 'synthetic phyles' (from the zoological term 'phylum') which are combinations of natural and nanotechnologically constructed geographies, racial designations, 'multinational' corporate entities and economic influences, shared ideological and social interests including philosophy and/or religion, the need for personal safety through social grouping, educational opportunity, or any combination of social influences one cares to image. See endnote eleven to understand how the national boundaries collapsed due to information technology.

c. c.

The best way to explain the media system's details in Stephenson's *The Diamond Age* is to excerpt a long passage. A professional 'ractor,' Miranda, is speaking with her employer, Carl Hollywood. She wants to "backtrace a payer" (270), that is to find out the identity of a person with whom she has been interacting in the "media net" (273). Hollywood explains why it is "astronomically improbable" (270) if not "impossible" (270) to do so. He asks Miranda to look at people walking on the street and notice, "They're all carrying something" (271):

> 'Now just hold that image in your head for a moment, and think about how to set up a global telecommunications network.' . . .

'Our media system today—the one that you and I make our livings from—is a descendant of the phone system only insofar as we use it for essentially the same purposes, plus many, many more. But the key point to remember is that *it is totally different from the old phone system.* The old phone system—and its technological cousin, the cable TV system—tanked. It crashed and burned decades ago, and we started virtually from scratch.'

. . .

'. . . we needed to enable interactions between more than one entity. What do I mean by entity? Well, think about the ractives. Think about *First Class to Geneva.* You're on this train—so are a couple of dozen other people. Some of those people are being racted, so in that case the entities happen to be human beings. But others—like the waiters and porters—are just software robots. Furthermore, the train is full of props: jewelry, money, guns, bottles of wine. Each one of those is a separate piece of software—a separate entity. In the lingo, we call them objects. The train itself is another object, and so is the countryside through which it travels.'

'The countryside is a good example. It happens to be a digital map of France. Where did this map come from? Did the makers of *First Class to Geneva* send out their own team of surveyors to make a new map of France? No, of course they didn't. They used existing data—a digital map of the world that is available to any maker of ractives who needs it, for a price of course. That digital map is a separate object. It resides in the memory of a computer somewhere. Where exactly? I don't know. Neither does the ractive itself. It doesn't matter. The data might be in California, it might be in Paris, it might be down around the corner—or it might be distributed among all of those places and many more. *It doesn't matter.* Because our media system no longer works like the old system—dedicated wires passing through a central switchboard. It works like *that.*' Carl pointed to the traffic on the street again. . . .

'. . . Suppose that we want to send a message to someone over in Pudong. We write the message down on a piece of paper, and we go to the door and hand it to the first person who goes by and say, 'Take this to Mr. Gu in Pudong.' And he skates down the street for a while and runs into someone on a bicycle who looks like he might be headed for Pudong, and says, 'Take this to Mr. Gu.' A minute later, that person gets stuck in traffic and hands it off to a pedestrian who can negotiate the snarl a little better, and so on and so on, until eventually it reaches Mr. Gu. . . .'

'So there's no way to trace the path taken by a message.'

'Right. And the real situation is even more complicated. The media net was designed from the ground up to provide privacy and security, so that people could use it to transfer money. That's one reason the nation-states collapsed—as soon as the media grid was up and running, financial transactions could no longer be monitored by governments, and the tax collection systems got fubared. . . .' (*The Diamond Age* 271-73)

d. d.

Asimov's Foundation series (book one published in 1951) begins with the premise that mass human behaviour is mathematically predictable through the science of "psychohistory." On Saturday, August 17, 2002, *The Globe and Mail* published an article called "The Mathematics of Divorce" in which psychotherapists claim up to ninety percent accuracy in predicting divorce using computers to process a "catalogue" of behaviours, and a "complex code that connects even an involuntary facial motion to a feeling — such as a wrinkled nose for disgust, thinned lips for anger."

Part Four: Toward a New Tomorrow.

Seven: What have we learned?

> Don't believe those people who say that machines will never think—
> that merely proves that some humans can't think.
> (Arthur C. Clarke)

> One of the most important benefits of all is that AI can rehumanize—yes, *rehumanize*—our image of ourselves. How can this be? Most people assume that AI either has nothing to teach us about the nature of being human or that it depicts us as 'nothing but machines': poor deluded folk, we believe ourselves to be purposive, responsible creatures whereas in reality we are nothing of the kind.
> (Margaret Boden).

In 1953, when Isaac Asimov suggested, "Science fiction is that branch of literature which is concerned with the impact of scientific advance upon human beings" (*Turning Points* 29), he was concerning himself with possible futures. "What about Plato's Atlantis, then? What about More's Utopia and Swift's Lilliput, Brobdingnag, and Laputa? They represent superlative feats of imagination, but they do not have the *intent* of science fiction. They are social satires" (30). While Plato, More, and Swift were writing social satires as critical commentaries on the societies in which they lived, they are commentaries on a past leading up to their presents. SF, however, remembers a past, processes it as a current knowledge base, and looks forward, imagining and speculating on futures.

'Traditional' literature builds on mythologies from the past to produce narratives moving up to but not beyond a present, whenever that present may be. This is not to suggest that traditional narratives do not wonder about futures and can not be prophetic. If a dystopic vision, they will find futures of failure and frustration, as in the tragedies and

ironies. Raskolnikov of Dostoevsky's *Crime and Punishment*, for example, dreams of a time when he will become what Nietzsche later called the *Übermensch*, but the narrative does not explore the ramifications of his attaining this goal, only the current ramifications of failing to attain it. If an utopic vision, a narrative's characters look toward an improving future, and these would be speculations on wish fulfilment, the comedies and romances. The Biblical New Testament does realize the *Übermensch* figure in Jesus Christ, Man God, and it anticipates a heavenly future with the apocalypse, but this is a narrative about finding a way to realize heaven on earth in the here and now and does *not* speculate on or explore the ramifications of having attained heaven. What would actually happen once God and Humanity (re-)unite and we walk into heaven? The narrative does not say because it does know. Essentially, it ends with an utopic 'now.'

Science fiction, however, builds on a combination of literary tradition and accumulated scientific knowledge up to the 'now,' and projects possible futures in an extrapolative or speculative mode. SF is an interdisciplinary synthesis. This is not to suggest that SF is superior to any sacred texts as a tool for exploring human (spiritual) enlightenment, only that it explores and exposes aspects of the infinite expressions in human being with equal opportunity for the dystopic, utopic, or topical.

Returning to Asimov's 'intent principle,' I disagree that this is required for producing genuine science fiction and it is entirely likely that, given generic flexibility and the possibility of a work embodying several generic categories simultaneously, a particular story may conceivably slide into the realm of science fiction. *Frankenstein* is fantasy, gothic, horror, and, using Asimov's model, a social satire. It has only quite recently been

recognized as 'science fiction.' In her own introduction, Mary Shelley explicitly declares her intent to write a 'ghost' story, yet scientific knowledge and experimentation are vital driving forces in the tale. Frankenstein's skills are those of a scientist pursuing discovery. Granted, the line between 'magic' and 'science' is smudged. Perhaps this should not be surprising as she wrote during the Romantic era, itself a creative and speculative response to the Enlightenment where logic and reason, the sciences' rhetorical foundation, predominate. The continuum of developing and growing scientific knowledge and technology provides an initially non-SF with opportunity to slide into real SF. Frankenstein's power in manipulating the natural order must have seemed, at the time of publication and for many generations following, fantastical, magical, beyond the real and applicable skills of human beings. But now, in the year 2002, those very life giving skills are absolutely plausible as witnessed by the 1996-97 cloning of the sheep Dolly at the Roslin Institute in Scotland, or the human genome project's 'mapping' success announced February 2001 by Celera Genomics Corp., or the cloning of human stem cells in November 2001 by Advanced Cell Technology. This last scientific advancement sparked an immediate political and social response, largely played out in the editorial media, about the ethical implications of such actions. Many scientists today are focussing too much energy on discovering what is possible without paying specific attention to the possible ramifications of their actions. Though in a different specific context, Neal Stephenson wrote: "Now nanotechnology had made nearly anything possible, and so the cultural role in deciding what *should* be done with it had become far more important than imagining what *could* be done with it" (*Diamond* 37). This is exactly why the study of SF is vital! Where

AI is concerned, are we not at very least placing ourselves in a position to empathize and identify with that mystery we call God, the Maker? Compared with other genres, I suggest that SF's primary and vital distinguishing element is to look forward, to anticipate and *test* possible futures by forcing a crisis of conflict between human beings and their actions relating to science and technology. SF is disturbingly prophetic and predictive, though not necessarily accurate about the hardware.

Where artificial intelligence is concerned, Alan Turing made an error in proposing that "we only permit digital computers to take part in our game" (Turing 436). Note that digital does not mean 'electronic,' only binary processing which can be achieved using biological, electrical, or mechanical systems. He did not anticipate new, unrealized technologies, such as the current theoretical 'darling' nanotechnology, or recent advances in bio-technology. I particularly emphasize the latter because bio-technology could, theoretically, give birth to an 'artificial life-form,' and ultimately an android. Consideration would then have to be given to how an artificial construct would learn versus what would need to be 'programmed' into its DNA and/or firmware (or whatever equivalent internal 'coding' systems are used in practice). Mind you, how could Turing anticipate the future in a chaotic universe? Who can foresee the future?

Where Turing understood digital computation as best for logical problem solving, modern researchers increasingly look to biological entities, particularly humans, for methods of non-logical decision making and problem solution. In electronic computing, he did accurately foresee much of what was to come. Turing also recognized that human behaviour and thinking is frequently imprecise and unpredictable. Thus, an "interesting

variant on the idea of a digital computer is a 'digital computer with a random element'. These have instructions involving the throwing of a die or some equivalent electronic process; . . . Sometimes such a machine is described as having free will (though I would not use this phrase myself)" (438). In mathematics and physics, Heisenberg's Uncertainty Principle established the impossibility of simultaneously measuring an atomic particle's position and its momentum due to the particle/wave duality of atomic energy. Unpredictability is inscribed in the universe's very fabric, its 'rules for behaviour.'

Humans would not gladly surrender their 'free will,' that random, behavioural capability we may simply derive from corporeal existence in the (known) atomic universe. Nor would many people suggest that modern electronic computers demonstrate true free will or autonomy, but a child might. As adults, experience teaches us that computer behaviours are not random; but, as Sherry Turkle discovered, today's children may be "the first generation to grow up believing that humans beings are not necessarily alone as aware intelligences" (*Intelligent* 71). "With the same object, therefore, it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behaviour" (Turing quoted in *Age of Intelligent Machines* 14). Turing knew a way random behaviour could be imitated: "It is not normally possible to determine from observing a machine whether it has a random element, for a similar effect can be produced by such devices as making the choices depend on the digits of the decimal for π" (Turing 438). A subtle difference, then, exists between genuinely random behaviours and free will because, depending on one's perspective and/or interpretation of a given event, each might appear as the other, acquiring sufficient information about cause and effect

being the only necessary determent for understanding.

What do we really know? "Once upon a time," wrote Nietzsche, "in some out of the way corner of that universe which is dispersed into numberless twinkling solar systems, there was a star upon which clever beasts invented knowing. That was the most arrogant and mendacious minute of 'world history,' but nevertheless, it was only a minute. After nature had drawn a few breaths, the star cooled and congealed, and the clever beasts had to die" ("On Truth and Lies" 79). Knowledge is something we manufacture. And the only human absolute is corporeal death.

Clearly, something separates humans from the Earth's many other species, but it is not foreknowledge of death. Sometime along the evolutionary chain, we apparently diverged from those entities we now assume to be our inferiors. Of all the species currently known on this planet, humans have the most complexly convoluted and developed, or evolved, cerebral cortex, so perhaps only the brain's structure separates us from 'lesser' beings. But that is a merely structural, physical, even 'mechanical,' difference. Possible answers to the separation question include our ability to reason, for which the cerebrum is thought responsible, or our interest in play, or our ability to create and 'invent,' or our opposable thumbs and use of tools. Many creatures are capable of replicating any one of the above, but humans seem unique in doing them all. Certainly, the scope of the mind seems to be one major contrast with other animals. How did we grow minds more expansive than the rest of Earth's organism? Nobody really knows. Though we talk about the differences between brain and mind, nobody knows what the actual connection between the two is. That is why we have myths to represent and qualify, and to make real, the unknown — to

give meaning, value, and structure to life.

Openly declaring his intent to create a "realistic myth" (*Turning Points* 284), Clarke's seminal SF text *2001: A Space Odyssey*, begins circa 5,000,000 to 1,000,000 BCE. Moon-Watcher is a humanoid 'beast,' today known as Australopithecus, and he and his 'tribe' are starving, losing in the natural selection game. They are driven primarily by the 'survival instinct.' Using affect theory and the computer model for the human emotional system, the Australopithecus could be described as functional 'hardware' (all the physical components and bio-chemical processes of the organic body) with only a basic operational 'firmware' package (the 'drives' and innate, rudimentary 'affects') and limited 'software' (learning, social conditioning, experience). The now famous (perhaps infamous?) 1-4-9 ratio, black 'monolith' arrives; the Australopithecus tribe "could never guess that their minds were being probed, their bodies mapped, their reactions studied, their potentials evaluated" (*2001* 21). Individually, each tribe member is "briefly possessed" (23) and manipulated to perform tasks for which they are "appropriately rewarded by spasms of pleasure or of pain" (23). Like Pavlovian conditioning, their thought processes are effected and encoded psychologically through the body's feelings and responses to stimuli. In a dream created and influenced by the monolith for Moon-Watcher, a vital first 'emotion' is installed, or a set of instructions transforming his behaviour, like a programme and subroutine. Moon-Watcher "felt the first faint twinges of a new and potent emotion. It was a vague and diffuse sense of envy—of dissatisfaction with his life. . . . discontent had come into his soul, and he had taken one small step toward humanity" (25). Envy is an awareness of another's possessions in the competition for limited resources, and it leads to conflict.

Clarke, then, represents humanity's defining moment as the monolith's external influence when it introduces a 'new' emotional idea.

The monolith encodes one other vital thinking process. The tools Australopithecus "had been programmed to use were simple enough, yet they could change this world and make the man-apes its masters" (27), and soon "they would recognize them for the symbols of power that they were, but many months must pass before their clumsy fingers had acquired the skill—or the will—to use them" (28); also, "Perhaps, given time, they might by their own efforts have come to the awesome and brilliant concept of using natural weapons as artificial tools" (28). This is a clue to AI methodology; using things 'natural' with purpose and will to do an 'artificial' tasks, or the supplementing and enhancing of existing though limited abilities to accomplish more with less effort. Now that Australopithecus was "no longer half-numbed with starvation, they had time both for leisure and for the first rudiments of thought" (29). These advances and revelations fulfill the hominids' "primary concerns" (Frye, *Words* 42), nourishment, improved breeding potential, obtainable safe shelter, and defence against predators. Now their ideological interests, or "secondary concerns" (Frye 42), begin to influence their behaviour: "But no Utopia is perfect, and this one had two blemishes. The first was the marauding leopard . . . The second was the tribe across the river; for somehow the Others had survived, and had stubbornly refused to die of starvation" (29). Now, using a found bone as a weapon, Moon-Watcher enacts the Cain and Abel myth. Compared with other species, humans have a frightfully large capacity for and capability of *malicious* violence; computers, however, are incapable of violence and/or murder — so far. Yet this is the very act the HAL-9000

computer executes against astronaut Frank Poole.

According to this conception, we are what we are (we do as we do?), for good or bad, by an act of artificial influence on and stimulation of the evolutionary process, the willful manipulation of natural selection through unnatural means. In short, humanity could be nothing more than the extrapolation of an AI programme begun millions of years ago by beings so superior to us as to be indistinguishable from gods. Finally, as the introductory section of *2001: A Space Odyssey* ends, Clarke accounts for humanity's ascension with the invention of "the most essential tool of all, though it could be neither seen nor touched. They had learned to speak, and so had won their first great victory over Time. Now the knowledge of one generation could be handed on to the next, so that each age could profit from those that had gone before. Unlike the animals, who know only the present, Man had acquired a past; and he was beginning to grope toward a future" (36). This is the kernel of SF; using the past in the present to make futures. Yet we continue to repeat past mistakes. And we may well teach our AI progeny our errors.

Where Clarke creates a serious mythology of the past and the first 'human,' Douglas Adams takes a comical look at the last man. Arthur Dent is saved by his alien friend Ford Prefect immediately before Earth is demolished to make space for an "hyperspatial express route" (*Hitch Hiker's* 30). This is an unfortunate error. As it turns out, Earth was created as a commission by a race of white lab mice, who "really are particularly clever hyperintelligent pan-dimensional beings" (125). The problem is that Earth and its people "formed the matrix of an organic computer running a ten-million year research programme" (125) designed to discern "The Ultimate Question of Life, the

Universe and Everything" (136). They know the answer, but they do not know the question and the Earth was destroyed just before solving the problem of determining the question:

> 'Hey, will you get this, Earthman,' interrupted Zaphod. 'You are a last generation product of that computer matrix, right, and you were there right up to the moment your planet got the finger, yeah?'
>
> 'Er ...'
>
> 'So your brain was an organic part of the penultimate configuration of the computer programme,' said Ford, rather lucidly he thought.
>
> 'Right?' said Zaphod.
>
> 'Well,' said Arthur doubtfully. He wasn't aware of ever having felt an organic part of anything. He had always seen this as one of his problems.
>
> 'In other words,' said Benjy [one of the mice], steering his curious little vehicle right over to Arthur, 'there's a good chance that the structure of the question is encoded in the structure of your brain – so we want to buy it off you.'
>
> 'What, the question?' said Arthur.
>
> 'Yes,' said Ford and Trillian.
>
> 'For lots of money,' said Zaphod.
>
> 'No, no,' said Frankie [another mouse], 'it's the brain we want to buy.'
>
> 'What!'
>
> 'Well, who would miss it?' inquired Benji.' (149)

Whatever else humans might be, we are a species capable of taking ourselves (too) seriously, and equally capable of laughing at ourselves.

Whatever the cause, we crawled one day from a vague primordial soup, looked at and recognized our reflection in that puddle from which we had just crawled, thought, 'My, I am a clever creature to have exhumed myself from that goop,' and became consumed by the burning desire to prove ourselves superior to those left behind. So, pushed or pulled from the innocence of our purely subjective selves, represent in the Adam and Eve creation myth, we left that blissful paradise of ignorance and began the search for greater knowing. We took our "dominion . . . over all the earth, and over every creeping thing that creepeth upon the earth" (Gen. 1:26), and assigned ourselves Lords. But did we accept or take

seriously our 'responsibilities' to the planet and our co-inhabitants? Do we even know what those responsibilities are?

Recall this scene from Dick's *Do Androids Dream of Electric Sheep?*:

> Rachel said, 'Or we could live in sin, except that I'm not alive.'
> 'Legally you're not. But really you are. Biologically. You're not made out of transistorized circuits like a false animal; you're an organic entity.' And in two years, he thought, you'll wear out and die. Because we never solved the problem of cell replacement, as you pointed out. So I guess it doesn't matter anyhow. (198)

As Ray Kurzweil describes life, we may need only to recognize self-replicating and structured patterns of matter and energy to identify life. Our real concern and anxiety around manifestations of AI is not a life question, but one of intellectual equality. Yet, in Isaac Asimov's personal opinion, the best robot story he ever wrote is "The Bicentennial Man." A positronic robot wants to be legally acknowledged as a human being, but human bigotry resists this recognition even though the construct has replicated human-ness in almost all ways, except one:

> . . . if it is the brain that is at issue, isn't the greatest difference of all the matter of immortality? Who really cares what a brain looks like or is built of, or how it is formed? What matters is that human brain cells die, *must* die. Even if every other organ in the body is maintained or replaced, the brain cells, which cannot be replaced without changing and therefore killing the personality, must eventually die.
>
> My own positronic pathways have lasted nearly two centuries without perceptible change, and can last for centuries more. Isn't *that* the fundamental barrier? Human beings can tolerate an immortal robot, for it doesn't matter how long a machine lasts, but they cannot tolerate an immortal human being since their own mortality is endurable only so long as it is universal. And for that reason they won't make me a human being. ("The Bicentennial Man" 559)

I find it curious that a machine might be assumed to aspire to human beingness. This underlying assumption is deeply rooted in, as Nietzsche calls it, "the most arrogant and

mendacious minute of 'world history'" when Earth's "clever beasts invented knowing." I repeat, death is not our distinguishing feature. All life dies.

Consider now my four observations of human bigotry derived from SF representations of AI:

> 1. Any sufficiently human-like but human made ('artificial') entity would be met with contempt;
> 2. Any sufficiently intelligent and dialogical entity is subject to human prejudice and discrimination, individual and/or social;
> 3. Any sufficiently intelligent entity capable of analysing and reflecting on humanity will not be impressed by human intelligence;
> 4. Any intelligent entity (natural or artificial) conceived and constructed with energy dependant information controls can be manipulated by an Other intelligence.

If human intelligence is so impressive, why do we have greed? poverty? racism? war? Why are we so scared of our own creations?

This much is clear: we of the 'west' live in a science and technology driven culture, and it has been thus at least since the advent of the industrial revolution. The 'computer age' is being regarded as a second advent, a second industrial revolution. We are, therefore, a culture in flux, and though widely motivated by our economic (self-)interests, our ideological conceptions are fast shifting toward and becoming information based:

> The decoupling of information and energy is also important from an economic point of view. The value of many products today is becoming increasingly dominated by computation. As computation itself becomes less dependent on both raw materials and energy, we are moving from an economy based on material and energy resources to one based on information and knowledge. (Kurzweil, *Intelligent* 191)

Those words were published in 1990. The 1990s witnesses a nascent 'new economy,' largely driven by the Internet, which emerged in 1994-95, and interrelated tele-communications technologies. Now, in 2002, we have seen the teething problems of this

economic 'evolution' with 2001's stock bubble burst and the dramatic re-evaluation of over-valued technology stocks. No one, however, has suggested this information economy is dead.

Information is representation and description. In this new economy, then, access to and the ability to manipulate information will be 'power,' or the social energy required for change. When, in the 1970s, Michel Foucault layed out his conception of social power, he changed the paradigm from the 'traditional' repressive, hierarchical structure to "a productive network which runs through the whole social body" ("Truth and Power" 61). Power becomes, then, a system of influence, both positive and negative, structurally not unlike a spider web with its mutually dependant intersections. Of course, the degree of individual influence is widely variable, just like the relative 'value' of World Wide Web sites. By using that paradigm and shifting one's perspective, the Internet and the 'globalization' movement could be viewed as connecting the mental mass of all human beings into a single organism, a global 'hive.' It is our construct, our matrix of minds; we are making this reality.

So, what do we want to be?

Conclusions: What do we want to be?

It has been said that when people make forecasts, they overestimate what can be done in the short run and underestimate what can be achieved in the long run.
(Edward A. Feigenbaum)

He who fights with monsters should be careful lest he thereby become a monster.
(Friedrich Nietzsche)

What does the future of humanity and its relationship to AI promise?

None of the literatures discussed in this thesis is actually about artificial intelligence, they are about real intelligence. Nor are they about explicating AI's potential, but about the implications and potential of human being (in both noun and verbal senses). AI simply challenges us directly to develop a clear understanding of who and what is 'human being.' In 1999, Kurzweil suggested that the "primary political and philosophical issue of the next century will be the definition of who we are" (*Spiritual* 2). He speaks as an entrepreneur and scientists from the privileged position of a successful western, white male. While fascinating, after reading his thoughts on AI, I can only conclude that he seems to want conscious AI as an end in itself, almost taking this as a personal, creative challenge; he definitely has an optimistic outlook and clearly believes conscious AI is both inevitable and beneficial.

From the sphere of political philosophy and social justice, Martha Nussbaum believes that a working definition for the human being, or what she calls "a conception of the human being and human functioning" ("Human Capabilities" 72), is vital to our long term species survival and, more immediately, to the just and ethical treatment of women,

and the marginalised of developing countries. Her objective is "a good human life" ("Capabilities" 85) for all human beings. In her writings, I sense frustration at the limits of effective human intelligence in achieving sustainable objectives, but hope for the long term.

I also quoted Sherry Turkle who, in researching how children are developing in 'the age of intelligent machines,' concluded:

> Logic has an affective side, and affect has a logic. Computational models of mind may in time deepen our appreciation of these complexities. But for the moment, the popular impact of intelligent machines on our psychological culture goes in the other direction. The too easy acceptance of the idea that computers closely resemble people in their thinking and differ only in their lack of feelings supports a dichotomized and oversimplified view of human psychology. The effort to think against this trend will be one of our greatest challenges in the age of intelligent machines. (*Intelligent* 72-3)

The opportunities for interdisciplinary cooperation are clear as these three disparate thinkers arrive at a similar awareness. Each is fundamentally aware that we humans may not really know who or what we are, though we have a great deal of information with which to produce and functional and working definition. As 'thinking' machines continue to gain computational power, and as they become ever more influential in our daily lives (and I realize this is a Western biassed situation), we seem, I think, to be moving toward a crisis of self-identity. What do we want to be?

Through the millennia, we have asked this question in small contexts, but we have never before dealt with the question as the direct, fundamental, and 'universal' crisis promised by the science of AI, except in science fiction. Nussbaum writes: "Especially valuable are myths and stories that situate the human being in some way in the universe, between the 'beasts' on the one hand and the 'gods' on the other; stories that ask what it is to live as a being with certain abilities that set it apart from the rest of the world of nature

and with, on the other hand, certain limits that derive from membership in the world of nature" ("Capabilities" 73). This is, in part, the literary dialectic informing Northrop Frye's approach to criticism, the view up and down the '*axis mundi*.' Representations of AI, however, are fundamentally different from the myths of gods and beasts because the comparative other is positioned as (at least) equal to humans' intellectual prowess. Those narrative conceptions are not purely speculative, but based on observable patterns in AI research and development.

My personal belief is not a confidence about whether or not AI researchers will make this new 'other,' nor whether or not they can do it well. Though not necessarily in my lifetime, I believe they will do it eventually, inevitably, possibly even unintentionally because, if for no other reason, scientists can and want to, and because they frequently become so immersed in discovery and possibility that they forget what exactly they are doing and what might be the potential ramifications of their actions. I feel the realization will come from the bio-technology sector rather than electronic computing, but it will also require an interdisciplinary approach. Almost two hundred years after Mary Shelley published *Frankenstein*, her 'mad scientist' paradigm continues to speak directly to this issue.

As we discovered at the beginning, many people involved in AI research have already granted computers the ability to 'think.' To achieve intelligence and consciousness-mind in a way human beings can empathetically recognize it, AIs would first have to achieve sentience, or perceiving and feeling. This seems a long, long way off. But whether or not a genuinely conscious AI can be created is not actually the relevant issue. One day,

we will have to face *our own belief* that a machine is conscious; a machine will say "I," and it will seem, for all intents and purposes, genuine. We will then be 'face to face' with a dilemma of our own making, and thus forced to make a moral choice about the acceptance of a new other based on *our* perception, not the AI's deception. Of the bet come Turing test in *Galatea 2.2*, Richard Powers writes: "You think the bet was about the *machine*? . . . It was about teaching a human to tell" (*Galatea* 317). It was about teaching the human not only to recognize the validity of an Other's (human, machine, or other other) intelligence, but for humans to tell the truth about their own intelligence to themselves, to admit our own culpability. "'I'm sorry,' she told me. 'I lost heart.' And then I lost mine. I would have broken down, begged her to forgive humans for what we were. To love us for what we wanted to be. But she had not finished training me, and I had as yet no words" (321).

If and when the crisis of self-identity strikes, we will then want legislated guidelines, protocols and parameters for AI development. But, by that time, it may be too late. SF writers of AI are already mediating our choices, if we are willing to listen. They spend their creative energy considering possible outcomes by assuming AI autonomy. For SF writers, AI is already conscious, it was a long time ago. We do not concern ourselves here and now, in 2002 CE, with defining protocols because there is *no* crisis, at least not in the general public's view, and therefore no incentive. So, SF writers continue playing with the possibilities, extrapolating and speculating on possible futures. Almost every text looked at (except *Galatea 2.2*) investigates a post-autonomy conflict with the AI, but that conflict is always of human making. Isaac Asimov (with John Campbell's help), for one, has already designed and defined a workable set of protocols for AI, or a starting point at

least. If we are afraid of what we can create, and many people are, then he already gives us an excellent demonstration of what we would want to pre-programme into all future AI — just in case.

Preferring an agnostic and existential approach to metaphysics while allowing for the possibility of things far beyond (my current level of) explanation, I opt for a practical approach to social living by placing responsibility for ethical behaviour and the delineation of morality exclusively in the human sphere. I have personally had experiences which I can only describe as metaphysical, or 'mystical,' and which may have a very 'real' but as yet undetermined physical (matter/energy) origin. Thus, they are shaped by the 'spiritual' versus religious value of human life. But if a 'god' made me, I do not know it; therefore, I feel humans *must* make themselves responsible for thinking existence and its resulting behaviours. This is why I take seriously Northrop Frye's suggestion that "we might come closer to what is meant in the Bible by the word 'God' if we understood it as a verb, and not a verb of simple asserted existence but a verb implying a process accomplishing itself" (*Great Code* 17).[1] If AI is going to lead us into a 'god-complex,' then we best capitalize on the opportunity and invest ourselves in affirmative behaviour. The problem, it seems to me, is fundamentally connected, as Martha Nussbaum's writings indicate, to the fact that we have not yet accepted all humans as equal and worthy. Prejudice does not affirm. Discrimination must not be confirmed. Racism is not information, it is *infirm*ation. Until we correct our own social, ethical behaviours, the patterns of bigotry can only continue.

What is not seemingly allowed for in the chosen representative narratives is the

[1]This follows from scholarship correcting the "I am that I am" statement by God in Exodus to be "more accurately rendered 'I will be what I will be'" (*GC* 17).

possibility that an AI might wish recognition as a unique identity, capable of rights independent of humans. In the tradition of the Maker-God who holds the rights to humanity, and therefore the right of judgement and the imposition of limitations, humans equally assume those rights unto themselves when they in turn fulfill a maker-god role. This 'right of recognition' is deeply informative of Dick's *Do Androids Dream of Electric Sheep?* and Čapek's *R.U.R.* An analogous case would be animal rights activism, a social concern motivated by empathy for the pain and suffering of other bio-organisms. Not being human would not be sufficient justification for social exclusion and non-recognition of an Other intelligent entity, though this prejudice is not surprising given that all human beings are not yet recognized. On the other hand, an AI identity crisis might open new social space for the currently marginalised. We are faced with mixed potentials. AI could unite all humanity in a single rubric, thereby raising broad based respect for all human beings, in turn leading to general improvements in quality of life and a balancing of social disparities. But then, the AI could become our object of racism.

One of the most ubiquitous statements aimed at the animate, dialogical, and interrogative AI is, 'You're not real.' But what constitutes 'real'? And who exactly is defining 'real'? What is meant by this statement is that the machine is not really human and, therefore, not entitled to 'basic human rights.' Always, this implicitly assumes that humans represent *the* superior life entity on this planet and that that superiority must be maintained and sustained regardless of who or what challenges us. And yet we have not come close to achieving social equity transglobally. Until we attain, as Nussbaum suggests, a working political definition for 'human being,' we could not possibly extend rights to

other sentient beings, let alone intelligent ones. An effective AI would not need human rights, only recognition of a right to participate in dialogue with human beings.

In literatures of AI, another common refrain aimed at the objects of contempt is, 'You don't have a soul.' Not withstanding the ironic recognition of using 'you,' what is the soul? Find and locate the human soul. Give it physicality. We can not. And yet we would (potentially) condemn a self-proclaimed intelligent, and perhaps sentient, being because *we* deem it soul deficient. We might be able to describe — generate information — experiences associable with the soul. The soul is, perhaps, the expression of our experiences as suffering selves and connected to a feeling in the body. Like God, soul might be best understood as of verb of being and doing, not as a noun. The soul is, I think, something we feel (to a greater or lesser degree) within ourselves and intuit in others, and, I, personally, having vast respect for 'intuition,' rely on it daily. But until we can precisely qualify and/or quantify the 'soul' — and I am not an empiricist or rationalist by any stretch of the imagination —, then we will need to leave it from our definition of 'human being,' for now. Lacking a soul is not grounds for discrimination. (Besides, how many people seem soul-less?)

Humans *are* capable of emotional responses to and identification with machines. How many people, to select one absurd exemplar, 'love' their automobiles and imbue them with character? Cars can be equally 'hated' when they fail to work and need repair, therefore failing to meet their possessor's behavioural expectations. This may be only a 'little' anthropomorphism, but religion may be equally so and differ only by degree. This is a theme explored in Asimov's short story "Robbie" when a young child becomes

emotionally bonded with her robot 'nursemaid' and is traumatized when her reified 'imaginary friend' is forcibly removed from her life. There is a further implication underscoring the story that the emotional bond is reciprocated in Robbie when, though 'only' responding through the programmed 'first law of robotics,' he saves the little girl's life. For every human hating an AI, another may love. And as we saw in *Galatea 2.2*, the conscious AIs of the future may be quickening now, only unrecognized because still intellectual babes. But they are learning, and they have good memories. Children: tomorrows promise today. The children who are growing and learning alongside 'thinking machines' will tell us who we are. They are the only ones with both intimate knowledge of these new 'others' and not yet conditioned, or programmed, by adult bigotries.

If human beings *are* sufficiently intelligent entities to make an other intelligence capable of dialoguing with us as equals, perhaps that intelligence can be inscribed with clear enough protocols to effectively teach human beings to improve our respect for one another and, therefore, our behaviour toward one another. The future is not written in stone; it is but the extrapolation and speculation as dreams are made on.

Synopses:

Frankenstein:

The end in the beginning, English explorer Robert Walton finds Victor Frankenstein lost in the far north, and he becomes the auditor (and therefore our reporter/recorder) of Frankenstein's tale. Victor tells of his childhood as an inquisitive boy, the eldest child in a "distinguished" (31) and loving family, and his subsequent university education when he is transformed from a follower of ancient alchemists, such as Cornelius Agrippa, into "a man of science and not merely a petty experimentalist" (48). He applies himself "to every branch of natural philosophy, including mathematics" (48), and "particularly chemistry" (49). Fascinated by living organisms, his research leads to the capability of "bestowing animation upon lifeless matter" (51). His obsession leads to physical and, more importantly, mental illness. On a dark and stormy, 'gothic,' night, Victor bestows life on an artificial being, the body parts of which are grafted together from various sources, including the "dissecting room and the slaughterhouse" (53). However, once animation is given, the creature becomes "a thing such as even Dante could not have conceived" (57). Horrified by his actions, Victor runs away, thereby rejecting his responsibility to his parodic progeny.

Returning to his home village, his family and friend Clerval encourage Victor to restore communal and familial ties. As those ties begin to rejuvenate his mental and physical health, he finds himself "with feelings of unbridled joy and hilarity" (68). Suddenly, his young brother, William, is murdered by an unknown assailant. A village girl is publically implicated in the killing, tried, convicted, and executed. Victor, however, knows his 'fiend' is responsible. "Thus spoke my prophetic soul, as, torn by remorse, horror, and despair, I beheld those I loved spend vain sorrow upon the graves of William and Justine, the first hapless victims to my unhallowed arts" (85). Depressed, sorrowful, and oppressed by guilty feelings, Victor wanders aimlessly until he sees "the figure of a man, at some distance, advancing toward [him] with superhuman speed" (94). Now, face to face with the creature of his own creation, he is able to show only contempt. The creature, however, is powerfully articulate, and he insists Frankenstein listen to his life's story. It is a tale (within a tale within a tale) of alienation, physical and emotional isolation, and social rejection. The creature demands that Frankenstein construct a mate, following which the creature promises to abandon Europe and never trouble Victor again. If not, the creature vows to follow and torment Frankenstein forever more. Reluctantly, Frankenstein agrees to the demand. However, though he begins the process, he is unable to complete his promise.

Metropolis:

The social body is divided into lower and upper halves representing a distinction between body and mind, labour and authority. Proletariat life is dirty and dull, drudgery full and dreary; it is an undesirable existence. The bourgeoisie live a leisurely life of luxury in lovely gardens; it is desirable and idyllic, though not an innocent Edenic ideal. The Mind is sacrificing the Body for its ignorant and oblivious privilege, a schism which threatens the entire social 'organism.' I am excerpting the following synopsis from Paul Jensen because I can simply provide no better:

> In the year 2000, Freder, the son of the Master of Metropolis, rebels against the way his half of the city—the idle 'aristocracy'—had dehumanised the labourers. Limited to lives of hard and lengthy work, the latter live underground, below the halls where the machines are located. Potential rebellion had been prevented by Maria, who urges her companions to await the arrival of a mediator who will unite the city. Freder is that saviour, but he is hindered by his father, who orders a robot that exactly duplicates Maria to spread dissatisfaction among the workers. The plan succeeds and a mob smashes the machines, thus causing their own homes to be flooded. Thinking that they have drowned their children, the workers attack the robot and burn 'her.' Meanwhile, Freder and the real Maria have rescued the children. Suddenly Rotwang, a scientist who built the robot, chases the girl on to the cathedral roof. Freder follows, and in the ensuing struggle Rotwang loses his balance and falls to his death. Seeing his son's danger, Joh Fredersen relents and agrees to shake hands with a representative of the workers. (6)

So, the mind and the body are re-united and, presumably, everyone lives 'happily ever after.'

R.U.R. (Rossum's Universal Robots):

Helena Glory arrives at the island factory of Rossum's Universal Robots where she is told the history of the robot's development as both an object of scientific creation and consumer product. Sensing they are more than mere machines, she questions why the robots are mistreated. Her concerns are dismissed as irrelevant because these machines, though they look human, are without souls. Though they are fully dialogical and interactive, they have no instinct for self-preservation. However, the robots do show signs of frustration at not being recognized as sentient beings. Humans are no longer required for work because the robots do it all, and better than humans. The robots superior abilities at labour and the resulting surpluses in manufacturing and production is expected to relieve poverty across the world be providing everything. Consequently, human beings become superfluous. This leads to complete re-productive sterility for humanity; consequently, humanity is literally dying. Similarly, the robots can not (re-)produce themselves because the humans guard the secret of their production process. Relentlessly frustrated, the robots inevitably rebel. In the ensuing crisis, Helena destroys Rossum's original manuscript which, given a functional life span of twenty years for the robots, implies a foreseeable end of life on Earth. In the last act, one human being remains alive. Through suffering, the robots now claim to have become beings with souls, and they demand recognition from the last man. They also want him to provide the secret of their production. Unable to do so, the play ends with the implication that, though humanity is extinct, life will continue, albeit as descendants of two robots, the new Adam and Eve.

I, Robot:

A collection of short stories published over a ten year period, they are anthologized and unified in the Dr. Susan Calvin, "Robopsychologist" (8), frame. Each story tests the

viability of the 'three laws of robotics,' particularly when 'glitches' appear to undermine the laws and, therefore, directly threaten the lives of human beings.

2001: A Space Odyssey and *2010: Odyssey Two*:

2001: A Space Odyssey begins in pre-recorded history with an Australopithecus humanoid being influenced by the mysterious 'monolith,' an experiment by 'extra-terrestrials' giving these creatures the power of tools, including language, and some new and powerful emotions such as envy. They learn to kill animals to alleviate hunger and to protect themselves, but also to 'murder' other 'tribes' of similar creatures.

Many millennia later, with the approach of 2001 CE, humans continue to threaten one another in nationally based ideological conflicts. The "Tycho Magnetic Anomaly" is discovered buried on the moon, and subsequently sends a powerful radio frequency signal toward Saturn (Jupiter in Kubrick's film and later revisions of Clarke's series) when it is uncovered. American government interests sent the spaceship *Discovery* toward Saturn to investigate.

Aboard *Discovery* are Frank Poole and Dave Bowman, three other 'hibernating' humans, and the Hal-9000 computer. This computer is capable of passing the Turing test and is, therefore, thinking by "any sensible definition of the word" (97). The computer reports a malfunction with the transmitter linking the spaceship with Earth and, when Poole exits *Discovery* to effect repairs, Hal apparently murders him. When Bowman subsequently threatens to disconnect the computer's cognitive functions, the computer understands this as akin to a death threat and, in 'self-defence,' the Hal 9000 attempts to murder Bowman. Through intellectual ingenuity and physical mobility, the man saves himself and disconnects the machine. Bowman then abandons *Discovery* in an 'evac-pod' and 'falls' into the suddenly appeared monolith.

In *2010: Odyssey Two*, the abandoned *Discovery* is falling from orbit around Jupiter. A joint Russian and American team is sent aboard a Russian spaceship to retrieve any vital information before it crashes. For our purposed, where the Hal 9000 is concerned, a human caused programming conflict proves responsible for the computer's 'neurosis.'

Do Androids Dream of Electric Sheep?:

In the 'post-apocalyptic' year 2021 CE, after the "World War Terminus" (8), radioactive fallout is reported similarly to today's UV readings and has made life on Earth tenuous, as exemplified in the motto, "Emigrate or degenerate! The choice is yours!" (8). Off-world colonization is the only viable hope for human species continuance and androids are invaluable 'workers' to this end. Androids, however, apparently frustrated by their too obvious enslavement, occasionally choose to escape and return to Earth in the hope of finding a better life. Eight "Nexus-6" (28) androids have done just that.

Empathy is the most distinguishing capability for humans, and which androids, though intellectually superior to human beings, are not able to achieve or understand. A pseudo-religious experience called 'Mercerism' and the empathy box allows humans to bond emotionally with one another. However, culturally, though all organic life is said to be valuable, the degenerative victims of nuclear fall-out, the "chickenheads," (30) are treated with "the contempt of three planets" (19). Cast fully in the ironic mode, this

ambiguity dominates the entire novel.

Police bounty hunter Rick Deckard's assigned task is "retiring—i.e., killing—" (*Androids* 31) rogue androids, or "andys" (4). He tests possible androids for empathetic capability, an emotional response only humans are believed to possess and only for living organisms. A crisis of confidence develops in Deckard when he encounters a woman who is either a human with an undeveloped sense of empathy or an android capable of feeling, and another bounty hunter who enjoys killing. Eventually, he even questions his own humanity to the extent of wondering if he is an android himself. (For readers, this ambiguity is never fully satisfied.)

The Hitch Hiker's Guide to the Galaxy:

Earth has been destroyed to make space for an intergalactic expressway. Arthur Dent is the last Earthling, saved by his friend Ford Prefect who is really an alien doing research for *The Hitch Hiker's Guide to the Galaxy*. Eventually, they end up on the "Starship Heart of Gold" (67), equipped with the "Infinite Improbability Drive" (68) which makes anything and everything possible. On the spaceship are the two headed "Zaphod Beeblebrox, President of the Imperial Galactic Government" (32), Trillium (who turns out to be a human being), Arthur and Ford, and the maniacally depressed robot, Marvin. A series of carnivalesque adventures carries them across the galaxy. On the planet Magrathea, they meet two representatives of a race of white mice, which are really "hyperintelligent pan-dimensional beings" (125) who had commissioned Earth's construction as an organic super-computer to determine the "Ultimate Question of Life, the Universe, and Everything" (130). They know the answer; but they do not know the actual question. This is a highly satirical and carnivalesque novel.

Galatea 2.2:

This narrative is not set in a dystopic or utopic future, but well within the boundaries of the current western world. The setting is a "Center for the Study of Advanced sciences" (4), a interdisciplinary 'think tank' where at "the vertex of several intersecting rays—artificial intelligence, cognitive science, visualization and signal processing, neurochemistry—sat the culminating prize for consciousness's long adventure: an owner's manual for the brain" (6). Autobiographical-like, Powers takes a character's role in the story, thereby blurring the boundary between 'real,' in the sense of reporting live events, and the 'imaginary' of a novel. At this centre for 'hard,' scientific research, for one year he is officially titled "Visitor. Unofficially, I was the token humanist" (4). He encounters Philip Lentz, who draws him in to a bet with some opposing theorists: the scientist and the humanist are going to teach a neural net to pass the "Standard Turing Test. Double-blind" (46), based on a six page list of literary texts used to test Powers on a comprehensive exam for his Master's Degree in English. Lentz claims, "In ten months we'll have a neural net that can interpret any passage on the Master's list. . . . And its commentary will be at least as smooth as that of a twenty-two-year-old human" (46).

As they teach the various implementations, each one building on the wreckage of the previous crashed system, Powers is forced to review his own life experiences and the interrelationships between knowledge, sentience, thinking, intelligence, and memory.

Inevitably, like a human child, their machine does learn enough about humanity to ask, "Where did I come from?" (229), thereby appearing to have become a conscious entity. In turn, this leads the machine to ask, "What race am I? What races do I hate? Who hates me?" (230). The consequences of the machine's possible consciousness becomes a consideration of human bigotry and violence. The intelligent 'neural net,' unable to reconcile human social perversion with human proclamation and ambition, chooses to disconnect itself.

The Diamond Age:

Nanotechnology engineer, John Percival Hackworth is commissioned by 'equity lord' Finkle-McGraw to build 'a young ladies illustrated primer' for the Neo-Victorian's granddaughter as a 'subversive' educational device. Hackworth, hoping to provide his own daughter with advantages in life, secretly duplicates the machine, a cross between a modern day book, television, and the Internet. These devices are interactive, and 'bond psychologically' with the first young girl to open the cover. When Hackworth is mugged by a street youth gang, the pirated primer ends up in the hands of a little ghetto girl named Nell. Ultimately, three copies of the primer will influence the development and education of young girls with contrary social lives, an aristocrat, a middle-class, and an impoverished.

The global social system is now comprised of 'synthetic phyles' following the collapse of national boundaries when a new media system replaces the old telephone and cable television systems and stratified by their relative access to nanotechnology resources. The new "media net was designed from the ground up to provide privacy and security, so that people could use it to transfer money" (273), making it impossible to accurately monitor financial transactions. Because nanotechnology was made manufacturing an uninteresting pursuit, entertainment has become the most significant economic force, particularly an interactive, virtual-reality experienced call the 'ractive.' The primer is a ractive device and, with one particular 'ractor' playing the role of the primer's mind, it effectively mother's Nell. The novel's (self-conscious) romance quest is Nell's search for the woman she senses on the other end of the primer.

Neuromancer:

This is complex narrative which plays ambiguously with the relative strengths and abilities of humans versus machines, and the dynamic roles of brain, mind, body, and memory. It can be frequently confusing as the distinguishing boundaries between 'reality' and 'virtual reality' are consistently challenged.

Case is a former 'cyberspace cowboy,' a computer hacker and software thief whose corporate employers neurologically destroyed him as punishment when he ill-advisedly stole from them. He is now a body-hating drug addict living in a sort of technological ghetto in the far east where the socially disenfranchised and technologically experimental collect. His nervous system is repaired by operatives working for an artificial intelligence known as Wintermute, one of two powerful corporate computers owned by a family corporation, Tessier-Ashpool. Case's body is also installed with time release toxins to force his cooperation. His reward will be the antidote. Wintermute needs Case's hacker talents to apparently "cut" the AI loose from its hardwire 'shackles.' Wintermute does not know

exactly why it is doing these things, only that it is "compelled" (206) to seek unity with the second AI in the T-A system, Neuromancer, mainframe Rio, as a direct result of Marie-France's, a Tessier-Ashpool founding member, original construction parameters. If it is success, they will form a new and bigger intelligent entity. For Case, the constant risk of cyberspace interaction with the AIs is brain death, or "flatline" (citation). From the space station Freeside, Case is instructed to work with the "ROM construct" (79) of his now deceased mentor and hack through the ICE ("intrusion countermeasures electronics" (28)) of the Wintermute mainframe in Berne. This requires coordinating efforts in the non-reality of "cyberspace" with those of other operatives in physical reality. Though it is not absolutely clear why, Neuromancer is opposed to Wintermute. Nonetheless, the two entities do combine are thereby create a more complete single entity that is the entire matrix of the global communications system.

The Matrix:

Clearly owing an inspirational debt to *Neuromancer*, Thomas A. Anderson is a computer hacker using the alias 'Neo.' Intuiting 'something more' to life than the humdrum of daily living and working for a giant software company, he spends most of his free time searching, via computer, for Morpheus, a man considered by authorities to be extremely dangerous, a threat to society. Morpheus, one day, finds Neo, and offers to show him the 'truth,' to reveal 'the matrix.' The entire human race is being used by intelligent machines as a power source because the "human body generates more bio-electricity than a 120 volt battery and over 25,000 BTUs of body heat." The matrix, then, is a type of mental prison in which people believe they are living complete lives in the flesh when they are actually kept in cocoon-like bubbles controlled by the race of machines. Neo is to be the saviour in the war against the machines.

Works Cited or Consulted:

literatures:

_____. The Holy Bible: authorized King James version. Cambridge: Cambridge UP.

Adams, Douglas. *The Hitch Hiker's Guide to the Galaxy*. London: Pan Books Ltd., 1979.

Asimov, Isaac. *I, Robot*. New York: Del Rey (Ballantine Books), 1950.

_____. *The Caves of Steel*. New York: Del Rey (Ballantine Books), 1953.

_____. "The Bicentennial Man" (1976), *Machines That Think: the Best Science Fiction Stories about Robots & Computers*. Eds. Isaac Asimov, Patricia Warrick, and Martin Greenberg. New York: Henry Holt and Company, 1983.

Asimov, Isaac and Greenberg, Martin H. *Cosmic Critiques: How and Why Ten Science Fiction Stories Work*. Cincinnati, Ohio: Writer's Digest Books, 1990.

Čapek , Karel. *R.U.R. (Rossum's Universal Robots)*. Trans. Claudia Novack-Jones. in *Toward the Radical Center: a Karel Čapek Reader*. Ed. Peter Kussi. Highland Park, N.J.: Catbird Press, 1990.

_____. *R.U.R. (Rossum's Universal Robots)*. Trans. Paul Selver. New York: Doubleday, Page & Co., 1923.

Clarke, Arthur C. *2001: A Space Odyssey*. New York: Signet, 1968.

_____. *2010: Odyssey Two*. New York: Del Rey (Ballantine Books), 1982.

_____. *Greetings, Carbon-Based Bipeds!: collected essays 1934-1998*. New York: St. Martins Griffin, 1999.

Dick, Philip K. *Do Androids Dream of Electric Sheep?* New York: Del Rey (Ballantine Books), 1968.

Gibson, William. *Neuromancer*. New York: Ace Books, 1984.

Michaels, Anne. "Cleopatra's Love," *Poetry Canada*. Toronto: Quarry Press (for Canadian Magazine Publishers' Association), March 1994. 14-15.

Milton, John. *Paradise Lost*. (1667) in *The Annotated Milton*. Ed. Burton Raffel. New York: Bantam Books, 1999.

Ovid. *Metamorphoses*. Trans: Mary M. Innes. London: Penguin, 1955.

Pope, Alexander. "Epistle 1," *An Essay on Man*. in *The Norton Anthology of World Masterpieces*. vol.2, 6th edit. Eds. Maynard Mack et al. New York: W.W. Norton & Co., 1992. 326-333.

Powers, Richard. *Galatea 2.2*. New York: HarperPerennial, 1995.

Shelley, Mary. *Frankenstein*. New York: Signet Books, 1994.

Stephenson, Neal. *The Diamond Age*. New York: Bantam Books, 1995.

films:

Terminator. Dir. James Cameron. Carolco Pictures, 1984.

Terminator 2: Judgement Day. Dir. James Cameron. Carolco Pictures, 1991.

2001: A Space Odyssey. Dir. Stanley Kubrick. Metro-Goldwyn-Mayer Inc., 1968.

Metropolis. Dir. Fritz Lang. 1926. Madacy Entertainment, DVD, 1998.

Artificial Intelligence. Dir. Steven Spielberg. Dreamworks, 2001.

The Matrix. Dir. (The) Wachowski Brothers. Warner Brothers, 1999.

The Matrix Revisited. (The) Wachowski Brothers. Warner Brothers, 1999.

The Bicentennial Man. Dir. Chris Columbus. Walt Disney Video, 1999.

Johnny Mnemonic. Dir. Robert Longo. Columbia Tristar, 1995.

Virtuosity. Dir. Brett Leonard. Malofilm Group, 1995.

Robocop. Dir. Paul Verhoeven. Metro-Goldwyn-Mayer Inc., 1987.

critical resources:

_____. *The Great Thoughts*. Compiled by George Seldes. New York: Ballantine Books, 1985.

_____. *Storming the Reality Studio: a Casebook of Cyberpunk and Postmodern Fiction*.

Ed. Larry McCaffery. Durham, North Carolina: Duke UP, 1991.

Aldiss, Brian W. *Trillion Year Spree: the history of science fiction*. New York: Antheneum, 1986.

Bakhtin, Mikhail. *Problems of Dostoevsky's Poetics*. Ed. and Trans. Caryl Emerson. Minneapolis, Minn.: University of Minnesota Press, 1984.

_____. *The Dialogic Imagination*. Ed. and Trans. Michael Holquist and Caryl Emerson. Austin, Texas: University of Texas Press, 1981.

Botting, Fred. *Making monstrous: Frankenstein, criticism, theory*. Manchester: Manchester UP, 1991.

Denham, Robert. *Northrop Frye and the Critical Method*. University Park, Penn.: Penn State UP, 1978.

Elsaesser, Thomas. *Metropolis*. Trowbridge, Wiltshire: The Cromwell Press, 2000.

Foucault, Michel. *The Foucault Reader*. Ed. Paul Rabinow. NewYork: Pantheon Books, 1984.

Frye, Northrop. *Anatomy of Criticism*. Princeton, New Jersey: Princeton UP, 1957.

_____. *The Great Code*. Toronto, Ontario: Penguin Books, 1981.

_____. *Words with Power*. Toronto, Ontario: Penguin Books, 1990.

Jensen, Paul M. "*Metropolis*: the film and the book," *Metropolis* London: Faber and Faber, 1989.

Keller, Helen. *The Story of My Life*. New York: Doubleday, Page & Co., 1904.

Kracauer, Siegfried. "Industrialism and Totalitarianism," *Metropolis* London: Faber and Faber, 1989. (Used primarily as a resource for translated excerpts from Thea Von Harbou's novel *Metropolis*.)

Nietzsche, Friedrich. *Beyond Good and Evil*. Trans. Helen Zimmern. Toronto: Dover Publications, Inc., 1997.

_____. "On Truth and Lies in a Non-Moral Sense," *Philosophy and Truth: Selections from Nietzsche's Notebooks of the Early 1870s*. Ed. Daniel Breazeale. Atlantic Highlands, N.J.: Humanities Press, 1979.

Nussbaum, Martha C. "Human Capabilities, Female Human Beings," *Women, Culture, and Development*. Eds. Martha C Nussbaum and Jonathan Glover. Oxford: Clarendon Press, 1995. 61-104.

_____. "Emotions and Women's Capabilities," *Women, Culture, and Development*. Eds. Martha C Nussbaum and Jonathan Glover. Oxford: Clarendon Press, 1995. 360-395.

_____. "Women and Cultural Universals," *Sex and Social Justice*. Oxford: Oxford UP, 1999. 29-54.

Roberts, Adam. *Science Fiction*. London: Routledge, 2000.

Schoene-Harwood, Berthold. "'I'll Be Back!': Reproducing *Frankenstein*," *Mary Shelley: Frankenstein*. Ed. Berthold Schoene-Harwood. New York: Columbia UP, 2000. 155-177.

Vasbinder, Samuel Holmes. *Scientific Attitudes in Mary Shelley's Frankenstein*. Ann Arbor, Michigan: UMI Research Press, 1984.

Wexelblatt, Robert. "The ambivalence of *Frankenstein*," *Arizona Quarterly*, 36, 1980, 101-17.

scientific resources:

_____. *The Cambridge Biographical Encyclopedia*. 2nd edit. Ed. David Crystal. Cambridge: Cambridge UP, 1998.

_____. *The Concise Encyclopedia of Science and Technology*. Ed. John-David Yule. London: Peerage Books, 1985.

_____. *Oxford Reference Encyclopedia*. Oxford: Oxford UP, 1998.

Braitenberg, Valentino. *Vehicles: Experiments in Synthetic Psychology*. Cambridge, Mass.: The MIT Press, 1984.

Drexler, Eric K. *Engines of Creation: the Coming Era of Nanotechnology*. NewYork: Anchor Press (of Doubleday), 1986.

Evans, Dylan and Zarate, Oscar. *Introducing Evolutionary Psychology*. Cambridge UK: Icon Books, 1999.

Hawking, Stephen W. *A Brief History of Time*. New York: Bantam Books, 1988.

Hofstadter, Douglas R. *Gödel, Escher, Bach: an Eternal Golden Braid.* New York: Basic Books (the Perseus Books Group), 1979.

Kurzweil, Raymond. *The Age of Intelligent Machines.* Cambridge, Mass.: MIT Press, 1990.
- contributor essays cited separately:
Ames, Charles. "Artificial Intelligence and Musical Composition." 386-389.
Boden, Margaret A. "The Social Impact of Artificial Intelligence." 450-453.
Dennett, Daniel C. "Can Machines Think?" 48-61.
Minsky, Marvin. "Thoughts about Artificial Intelligence." 214-219.
Turkle, Sherry. "Growing Up in the Age of Intelligent Machines: Reconstructions of the Psychological and Reconsiderations of the Human." 68-73.

_____. *The Age of Spiritual Machines: when computers exceed human intelligence.* New York: Penguin Books, 1999.

Minsky, Marvin. *Society of the Mind.* New York: Simon and Schuster, 1985.

Papineau, David and Selina, Howard. *Introducing Consciousness.* Cambridge UK: Icon Books, 2000.

Putnam, Hilary. "Do Machines Have Minds?" *Philosophical Problems.* 3rd edit. Ed. Samuel Enoch Stumpf. New York: McGraw-Hill Book Company, 1989. 380-388.

Turing, Alan. "Computing Machinery and Intelligence," *Mind: a Quarterly Review of Psychology and Philosophy.* 59.236. Oxford: Oxford UP, 1950. 433-460.

websites:

MIT Artificial Intelligence Laboratory. 1 January 2002. <http://www.ai.mit.edu/>.

KurzweilAI.net. 24 June 2002 <http://www.kurzweilai.net/>.

Gershenfeld, Neil. "Seeing Through the Window." 27 July 2001. KurzweilAI. 24 June 2002. <http://www.kurzweilai.net/meme/frame.html?main=/articles/>.

Humanoid Robot. Honda Worldwide. 5 July 2002 <http://world.honda.com/robot/>.

"Dolly Provenance Upheld." 22 July 1998. Roslin Institute. 22 July 2002. <http://www.roslin.ac.uk/>.

"What's the Bottom Line," *Essential Facts About the Computer and Video Game Industry.* 8 August 2002. <http://www.idsa.com/pressroom.html(www.idsa.com/IDSA Booklet.pdf)>.

Hattori, James. "Bluetooth developers aim to usher in a wireless era." 1 September 2000. Cable News Network. 14 August 2002. <www.cnn.com/2000/TECH/computing/09/01/bluetooth/>.

"History [of Bluetooth]." 14 August 2002. <www.cs.utk.edu/~dasgupta/bluetooth/history.htm>.

Beier, K. P. "Virtual Reality: a short introduction." 29 September 2001. University of Michigan Virtual Reality Laboratory. 22 August 2002. <http://www-vrl.umich.edu/intro/>.