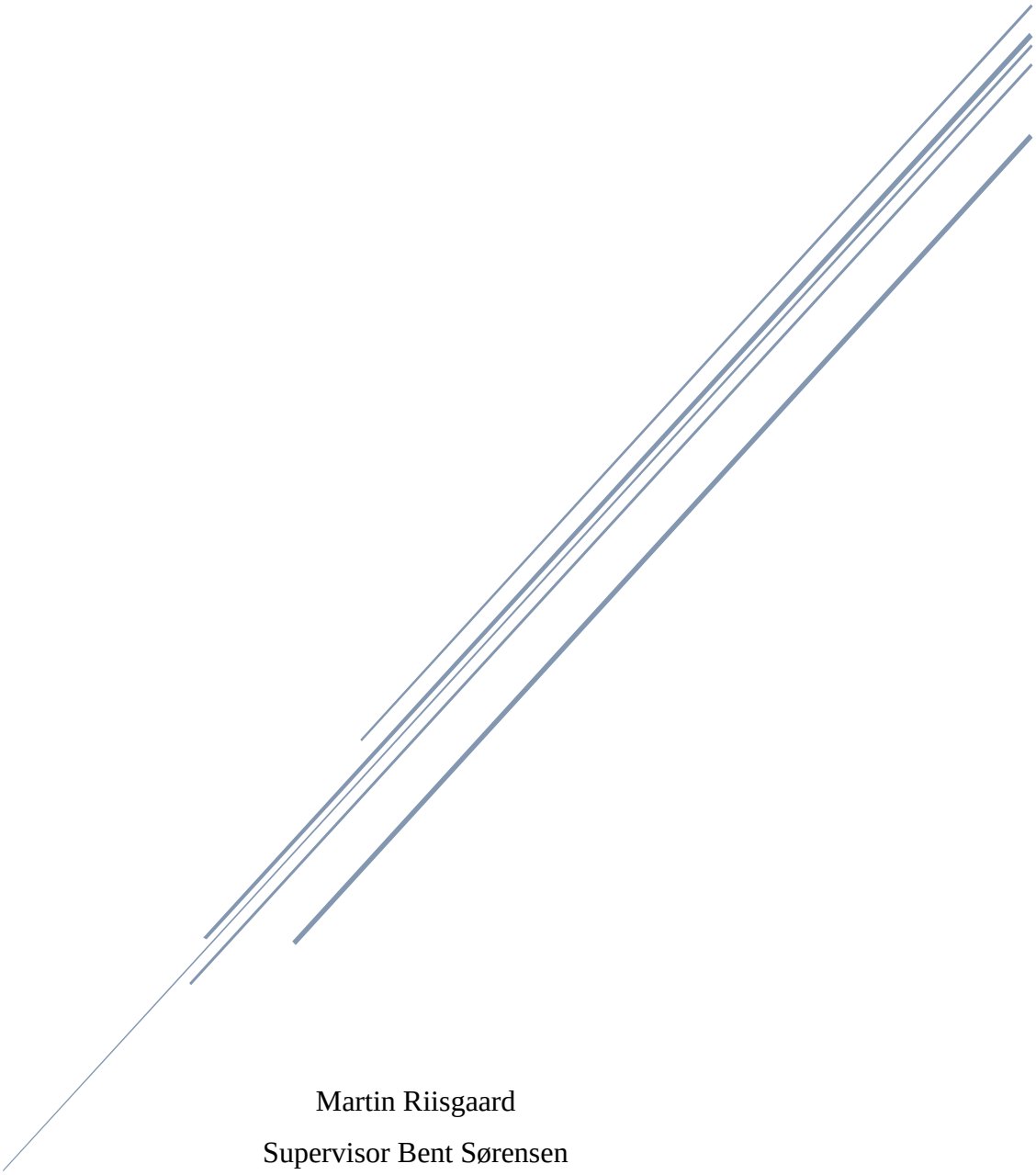
RAGE AGAINST THE MACHINES:

A comparative posthumanist analysis of interrelations between humanity and sentient machines



Martin Riisgaard

Supervisor Bent Sørensen

Master's Thesis in English

Aalborg University

Abstract

In a world that is constantly getting bombarded with media featuring menacing synthetic life, what tools can we use to understand and hopefully prepare ourselves to live side-by-side with synthetic life in the future? The thesis attempts to answer the following questions: *How is the interrelation between humanity and sentient machines presented in Banks'* Consider Phlebas *and Rucker's* Software*; how can posthumanism specifically be used to analyse this; and finally, how do these two texts approach the question of humanity and synthetic life co-existing?*

This thesis seeks to examine the interrelations between humanity and sentient machines in the two science fiction texts *Software* (1982) by Rudy Rucker and *Consider Phlebas* (1987) by Iain M. Banks. The examination and analysis of these two texts is undertaken through a close reading utilising a posthumanist theoretical framework based on N. Katherine Hayles' *How We Became Posthuman* along with considerations on the notion of 'personhood' from Anne Foerst. The paper finds that while the two texts are on a surface level quite similar, the strife between humanity and sentient machines predominantly arise from transgressing on the inherently inviolable nature of the *body/embodiment* polarity. Likewise, a perceived importance of consciousness was found to be adverse, if the humans and synthetic life is to thrive with one another. Unfortunately, the discussion on consciousness suffers from the fact that consciousness as a field is inherently to complex to accurately treat in this paper.

Contents

Introduction

Virtually since the birth of the first modern computers, the topic of intelligent machines and later artificial intelligence, AI, has been a hotly contested topic. Persistently present in this topic has been the British mathematician, Alan Turing, by posing the seemingly simple question whether machines "can think" in his influential paper "Computing Machinery and Intelligence" from 1950 (Turing 443). Among other thoughts concerning thinking machines, Turing introduced the Imitation Game, also frequently referred to as the Turing Test, which proposed a test structured as a conversation wherein the goal of the machine was to trick the human participants to such a degree that they would be unable to distinguish between the answers provided by the human and the machine, thereby seemingly demonstrating an intelligence similarly to that of a human being (Flasiński 3). Although, the usefulness of this test in contemporary sciences is doubtful, the test continues to haunt even present-day media, one such example being Alex Garland's 2014 science-fiction film *Ex Machina*.

Although the idea of truly intelligent machines, who are indistinguishable from humans, appears to reside somewhat in the far future, robots, androids, automatons and varies other types of machine intelligences are a truly ubiquitous theme within the realm of science-fiction texts. Of particular interest in this thesis is the meeting between humanity and sentient machines – the confrontation between biological- and artificial life. There are numerous examples in literature of this interaction both being fruitful but just as often having disastrous results: it can be argued that the monster from Mary Shelley's *Frankenstein; or, the Modern Prometheus* (1818) is one of the earliest examples of a sentient, artificial life in literature. However, the monster is perhaps more of a cyborg than an actual synthetic creation through and through. Another persistent player in the field of synthetic life is Isaac Asimov and his Three Laws of Robotics:

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

The legacy of these laws is so pervasive that the three laws are even referred to as existing in-universe in Rudy Rucker's *Ware Tetralogy* of which the first novel, *Software* is of particular interest to this thesis. To counterbalance the above three laws, which leaves little room for

interpretation that robots are slaves in all but name in Asimov's *I, Robot* (1950) universe, several films, video game and literature does provide amble proof that such laws might be necessary: films like James Cameron's *Terminator* (1984), the Wachowski siblings' *Matrix* (1999) and the aforementioned *Ex Machina* (2014) all feature homicidal sentient machines. Games like Bioware's *Mass Effect* (2007) and LookingGlass Technologies' *System Shock* (1994) also follow the trend of having sentient machines adverse to humanity's survival. Finally, novels like Philip K. Dick's *Do Androids Dream of Electric Sheep?* (1968) and William Gibson's *Neuromancer* (1984) likewise feature humans' and sentient machines' troubled relationship. The area of interest in this thesis will therefore seek to examine these relations – particularly in the two texts *Consider Phlebas* (1987) by Iain M. Banks and *Software* (1982) by Rudy Rucker. To provide the theoretical framework to study these texts in depth I will employ posthumanism to analyse the interactions between sentient machines and humans, and hopefully tease out what causes this discord between humanity and their synthetic counterparts.

Therefore, the thesis statement posed will be the following:

- *How is the interrelation between humanity and sentient machines presented in Banks'* Consider Phlebas *and Rucker's* Software*; how can posthumanism specifically be used to analyse this; and finally, how do these two texts approach the question of humanity and synthetic life co-existing?*

These titles are quite dissimilar in several aspects which will become apparent, but a comparison between the two should help demonstrate just how diverse literature has considered to the topic of humanity's meeting with intelligent life, and how it has shaped both. The choice of texts will be elaborated on in the 'methodology' section of the thesis, in which the choice of theoretical framework and finally data collection will also be discussed. Then I will account for posthumanism, particularly through N. Katherine Hayles' book *How We Became Posthuman*. After the 'theory' section follows first an analysis of *Consider Phlebas* and then *Software*. Finally, I discuss the findings from the analysis with posthumanism. I then proceed to conclude the findings of the thesis and discuss other possible approaches along with weaknesses of the thesis.

Before moving on to methodology, I would like to bring attention to a convoluted problem in this thesis: namely how to refer to intelligent machines consistently and most accurately? Already, expressions like Artificial Intelligence (AI), intelligent machines, sentient machines have been used. Later, I will also use the expressions synthetic life and robots. Furthermore, like AI there is also a field of study called AL, Artificial Life. While I

use most of these expressions interchangeably, the favoured expression in this thesis is sentient machines as that one seems to be the most neutral of the expressions. However, I do not have a sufficiently knowledgeable background to give a reasonable assessment of which term is the best. Likewise, I am not going to go into a deeper exploration of what exactly constitutes 'sentience' or 'intelligence' in this thesis as that also lies well outside the scope of this thesis. But as we shall see, even the word machine could prove problematic, so I will just leave this disclaimer here.

Methodology

As the introduction made clear, the goal for this thesis is to examine humans' and intelligent artificial life's interaction with each other. In this part of the thesis, I will attempt to explain some of my reasonings for the choices that I made in trying to achieve this goal. Specifically, I will discuss the two novels here, and likewise mention some other novels that I considered using while also giving explanations for why I did not utilise them. I will also account for the data that I obtained in the paper, and what my steps were to procure this data. Finally, I will also discuss the choice of theoretical approach including the literature in this part of the paper.

The most straightforward approach to begin to explain the choice of texts in this thesis is likely by talking about genre briefly. If one were to boil down my thesis statement, it is accurate to say it deals with 'human and intelligent synthetic Other interacting'. In this case, I am limiting myself to the more *speculative fiction* literary genres since humanity is arguably quite alone in that regard, as it is right now in our reality. Obviously, speculative fiction genres provide a plethora of possible avenues to examine humanness/Otherness interactions from elves, to dwarves, to ghosts or to aliens there are a myriad of options. Specifically, choosing to focus on artificial intelligence or sentient machines is a choice that has been shaped by previous papers treating William Gibson's *Neuromancer* (1984) and the tv series *Person of Interest* (2011), which both feature a hyper intelligent AI that incite fear because of its existence. Continuing to pursue this interest seemed obvious. The choice of texts, explicitly, is a combination of the following factors: genre, publication date, plot and interest.
Genre: *Software* is a cyberpunk novel while *Consider Phlebas* is a space opera. With my previous experience reading cyberpunk (Neuromancer), I chose to diversify with the second text to hopefully find interesting divergences.
Publication date: *Software* was published 1982 and *Consider Phlebas* was published 1987. Initially, I considered using Isaac Asimov's short story collection *I, Robot* (1950), but it

seemed that the gap between it, and the other novels from the 1980's would lead to a different analysis and discussion.

Plot: From my previous work with *Neuromancer*, it was evident to me that the texts needed to have sentient machines as a major contributor to the plot. Just as necessary was that the machines had to have thoughts and feelings that were implicitly and explicitly stated by themselves.

Both texts chosen are award-winning novels, which naturally also piqued my interest.

Data

As the theory concerns posthumanism and the point of the thesis itself is to examine human and artificial sentience relations, the chief focus has been on posthuman aspects of the 'humans' of the novels and the sentient machines themselves. Since this is a qualitative study, the approach to collect data has been *close reading* of the texts with the two topics: posthuman humans and sentient machines. Now, close reading by itself presents a bit of an issue for a potential reproduction of my results. Matthew Jockers highlights the problems of the results stemming from close reading to that of 'science': "The conclusions we reach as literary scholars are rarely 'testable' in the way that scientific conclusions are testable. And the conclusions we reach as literary scholars are rarely 'repeatable' in the way that scientific experiments are repeatable. We are highly invested in interpretations, and it is very difficult to 'rule out' an interpretation" (6). However, as Jockers then goes on to point out, the interpretations made by one scholar, or even layman, can lead another to "see" something in the text that they previously could not, thus leading to a greater understanding of the text (6). Finally, Jockers emphasises that interpretation is the product of observations, and observations are "both in the sciences and in the humanities" flawed (6). For even if the results are indisputable, the observers are always subject to bias. To minimise my bias in my close reading of *Consider Phlebas* and *Software*, I initially made it a point to make a note every time an artificial entity was mentioned. Later, I also realised that I had to, likewise, take a closer look at the posthuman-humans than I had initially planned. One bias that I would like to admit at this point is that I am personally optimistic about a potential future of human-sentient machine relations.

Theoretical Approach

The theoretical approach has already been mentioned a few times as posthuman, but this paragraph will be devoted to elaborating on the choices and thoughts behind the exact theoretical approach. While I choose to use examine posthumanism primarily through the lens of N. Katherine Hayles and her book *How We Became Posthuman*, I will first account

for another principal text within the posthumanist field, and why I chose to forego talking about it: Donna Haraway's *A Cyborg Manifesto*. First, while *A Cyborg Manifesto* (hereafter referred to simply as 'cyborg manifesto') certainly encapsulates posthuman thoughts in that the boundaries between human-animal-machine are explored, Haraway's cyborg appears to be less than ideal for talking about beings that are completely artificial: "She offered the cyborg as a contemporary cultural metaphor in order to capture the ambivalent condition of the contemporary human beings, whose bodies are open to forms of technological modification and intervention" (Bolter 2). Additionally, the cyborg manifesto remains a significant essay within feminism as well. However, this thesis will not directly involve itself with feminist theory or thoughts. As the theory section elaborates on posthumanism, it will be clear that it is by its nature of opposing "humanism itself postulated or simply assumed man as the standard and norm of the human" closely related to feminist works that seek to "[displace] man from his central position in the definition of the human and challenges the coherence of gender as a category (Bolter 4).

Hayles' *How We Became Posthuman* diverges to an extent from Haraway's approach as Hayles' approaches posthumanism through a foundation in cybernetics and informatics. Early cybernetics, as we shall see, tries to: "ultimately […] describe the human subject as a mechanism" (Bolter 4), which almost seems like an antithetical approach to defining human beings compared to the more humanist viewpoint that posthumanism is trying to break with. The notion of embodiment is also central for Hayles with consciousness being of lesser, or even no, import. These and other considerations are why a predominant part of the theory in this thesis follows Hayles closely.

Lastly, the field of transhumanism is often mentioned in conjunction with posthumanism, and Hayles, likewise, uses Hans Moravec *Mind Children* to emphasise a nightmare-like scenario future scenario for humanity that transhumanists like Moravec eagerly await. For this reason and to be able to talk about more critically about Rucker's *Software*, there is a short section on Moravec's transhumanism in the theory, although it will not be part of the analysis and discussion other than to serve as small contrast.

Theory

This section of the thesis will delve into posthumanism which will be used later in to discuss the findings of the analysis to tie them into a greater whole, subsequently form the basis for a discussion that reflects the thesis statement. The structure of this section will be as follows: A short introduction to the influence of cybernetics on Hayles' definition of posthumanism. I will go through cybernetics relatively hastily with a chronological overview of each

order/wave of cybernetics, after which I'll have a paragraph on the liberal humanist subject. This is important as the posthumanist being is distinctly related to the humanist subject. Then I proceed to account for Hayles' posthumanist framework – particularly that of the body/embodiment duality and the question of consciousness. I employ Anne Foerst to further establish a framework for discussing the notion of "personhood". Finally, the chapter ends with a short accounting of *transhumanism* through Hans Moravec, and subsequently Charles T. Rubin's refuting of transhumanism as a functional future for humanity.

Cybernetics

In *How we became posthuman* Hayles considers the field of cybernetics as a formative component of posthumanism, which also leads to this section serving as an overview to the field of cybernetics. Generally speaking, cybernetics is, according to D.A. Novikov, "the science of general regularities of *control* and *information transmission* processes in different systems, whether *machines*, *animals* or *society*" (1; emphasis in original). Cybernetics is an interdisciplinary science which includes fields such as mathematics, logic, semiotics, physiology, biology and sociology with area of study being "concepts of control and communication in living organisms, machine and organizations including self-organization" (Novikov 1). According to Novikov the "modern (and classical!) interpretation of [cybernetics]" can be traced to Norbert Wiener who defined cybernetics as: "the scientific study of control and communication in the animal and the machine", later adding society as the third object of interest for cybernetics (2). According to Hayles, the emerging interdisciplinary field of cybernetics was no less than a paradigm shift with the result being that "humans were to be seen primarily as information-processing entities who are *essentially* similar to intelligent machines" (7; emphasis in original). As is natural for a discipline that has been an area of active study for 50+ years, cybernetics has developed and expanded beyond Wiener's original definition.

One such development is the second-order cybernetics in which the central idea is that of an *observer*, which is still an integral part of any activity, domain or model. Novikov gives this definition of the phenomenon: "The cybernetician, by entering his own domain, has to account for his or her own activity" (10). Less accepted are the notion of "third-order" - and even "fourth-order" cybernetics (Novikov 11). Hayles herself seems to accept third-order cybernetics so a short definition will be presented here: "third-order cybernetics can be formed basing on the thesis 'from observing systems to self-developing systems'" (V. Lepsky, qtd. in Novikov 11). Hayles replaces the word *order* with *wave* and for simplicity's

sake this will likewise be adopted here. Although this is a short overview, the following paragraphs will delve deeper into the different cybernetic waves, as they become relevant.

1st Wave Cybernetics

In the line of investigation into how information got disembodied, the Macy Conferences were instrumental in pinpoiting the shift whereby information became "a disembodied medium" (Hayles 50). Initially Norbert Wiener and John von Neumann paved the way for information's importance in Cybernetics by "making clear that the important entity in the man-machine equation was information, not energy" (Hayles 51). Specifically, with information the importance was on the ability for it to flow through a system and how rapidly it could do it. Wiener presents the idea explicitly as: "The fundamental idea is the message … and the fundamental element of the message is the decision" (qtd. in Hayles 52), meaning that the decision is what produces information in the end. Taken to its extreme Wiener and Claude Shannon saw information as representing a choice as the quote from Wiener above certainly illustrates. Specifically, the idea is that the choice of one message represents one choice out of many other 'not-chosen' messages. Importantly, both Wiener and Shannon sought to distance information from meaning since this would allow it to have a *stable value*. Having information still attached to meaning, would change its value as it was used in different contexts, as "context affects meaning" (Hayles 53). To illustrate this, Hayles used the example of the sentence '*it's raining'*. If you are standing inside a windowless building, the meaning is different from saying the same sentence while standing outside in the rain, which presumably would be seen as ironic as opposed to the first scenario which might serve a prudent warning (Hayles 53). Strictly speaking, the information provided is the same, but the context changes the meaning of the message. This approach to thinking about information can be summarised as "what it *is*" as opposed to other approaches to define "what it *does*" (Hayles 56). These distinctly diverse outlooks mean that the Shannon/Wiener theory, by stripping the context from information, successfully renders it a calculable mathematical quantity - free from the bounds of materiality and indeed bodies (Hayles 56). Finally, we arrive at information as a disembodied entity. Opposite this, the "what it does" approach sees information more as an action than a reified object. These two opposing views on the 'problem' of information further illustrate the divide between first wave and second wave cybernetics with their focus on homeostasis and reflexivity respectively (Hayles 56). Seeing information as a quantifiable 'thing' enables a homeostatic perspective as it can generally be transferred freely, while information as an action necessitates an awareness of what the effect on the receiver conceivably might be (Hayles 57). Having reached one of the

conflicting areas of interest between the first – and second wave cybernetics, it is only instinctive to move on to an account of this movement.

2nd Wave Cybernetics

As was briefly remarked on, the second wave of cybernetics turned "the focus from the observation back onto the observer" (Hayles 134), but before delving further into this assertation, it is prudent to give a small account of the unassuming creature that came to symbolise this movement in cybernetics: the modest frog. Several prominent Macy Conference members, J. Y. Lettvin, H. R. Maturana, W. S. McCulloch and W. H. Pitts, made contributions to the article titled: "What the Frog's Eye Tells the Frog's Brain", whose findings demonstrated "that the frog's visual system does not so much *represent* reality as *construct* it" (Hayles 131; emphasis in original; see Lettvin, et al.: *What the Frog's Eye Tells the Frog's Brain*). The common assumption was then that what holds true for frogs, likewise, holds true, neurologically, for humans. There is no reason to assume that humans are uniquely equipped to perceive the world "as it 'really' is" as Hayles puts it (131). The primary driving forces in the second wave cybernetics, which concerned itself with "talking about life and about the observer's role in describing living systems" (Hayles 131), was Chilean neurophysiologist Humberto Maturana and his collaborator Francisco Varela.

The addition of reflexivity in cybernetics was started by Austrian immigrant Heinz von Foerster. Although he himself had served as principal editor of the Macy Conferences' transcripts, it was not until after they had concluded that he attempted to expand on the "epistemological implications of including the observer as part of the system" (Hayles 133). One problem facing reflexivity in its ascendency was the issue of shielding it from the realm of subjectivity. With its mathematical grounding, cybernetics requires a definition that distinctly moves away from subjectivity. In 1969 Foerster encounters Maturana's ideas "about treating 'cognition as a biological phenomenon'" (qtd. in Hayles 134), which affected Foerster and subsequently helped him formulate reflexivity in a distinctly cybernetically useful style: "Instead of searching for mechanisms in the environment that turn organisms into trivial machines, we have to find the mechanisms within the organisms that enable them to turn their environment into a trivial machine" (qtd. in Hayles 134). These findings seem to echo the conclusions drawn from the article "What the Frog's Eye Tells the Frog's Brain" wherein researchers could see a frog's reaction to certain different stimuli. What became apparent to them, was that the frog's brain emphasised reacting to fast moving objects over slower moving bigger objects, which seems natural given the frog's natural interest in insects and disinterest in bigger animals/stimuli. What this specifically indicates, is that the frog does

not transmit the 'precise' image its eyes perceive, but rather a form of organised interpretation of what it sees. As Maturana and other scholars continued their research into colour vision in other animals, they soon discovered that there was no part of the nervous system that colour vision was specifically mapped onto: "there was no one-to-one correlation between perception and the world" (Hayles 135). This and other studies lead Maturana to the conclusion that "perception is not fundamentally representational", therefore he further argued that it is not possible to speak of an "objectively existing world" as the idea of a 'world' is a construction by an observer. What is commonly referred to as 'reality' only "comes into existence for us, and for all living creatures, *only through interactive processes determined solely by the organism's own organization*" (Hayles 136; emphasis in original). Maturana and Varela summarise it as such:

> No description of an absolute reality is possible. Such a description would require an interaction with the absolute to be described, but the representation which would arise from such an interaction would necessarily be determined by the autopoietic organization of the observer, not by the deforming agent; hence, the cognitive reality that it would generate would unavoidable be relative to the knower. (Maturana & Varela 121)

With the obvious importance on what or who Maturana's 'observer' is, it is evident that some more attention be devoted to this phenomenon.

It is paramount for Maturana that the observer is always considered: "*The observer is a living system and any understanding of cognition as a biological phenomenon must account for the observer and his role in it*" (Maturana & Varela 48; emphasis in original). The observation itself necessitates a certain amount of reflexivity from the observer as one of the systems that are always part of their observation is themselves/itself. Maturana's observer is explicitly conceptualised as simply the 'observer' without a "psychological depth or specificity" as Hayles puts it (143). The point is simply that the species of 'the observer' should not matter, as any observing individual performing the same observation should perceive more or less the same as any other observer (Hayles 143). Maturana's observer and reflexivity are closely intertwined as each process of observing a system is part of a reflexive circle: "as when an observer thinks, 'I am an observing system observing itself observing'" (Hayles 144). What this process does, is distinguishing second-wave cybernetics from first-wave cybernetics in terms of how information itself is viewed. While first-wave cybernetics attempts to partition information from its body, Maturana's autopoietic theory "draws attention to the fact that 'information', so defined, is an abstraction that has no basis in the

physically embodied processes constituting all living entities" (Hayles 149). In other words, information without a body cannot exist as first-wave cybernetics supposes. Now, we will explore the next wave of cybernetics.

3rd Wave Cybernetics

As opposed to second-wave cybernetics, which trended towards circular processes, third-wave cybernetics' processes are best described as being spiral in nature. Whereas the inclusion of the observer is paramount in second-wave cybernetics, the focal point in third-wave cybernetics is to a greater extent on how to develop systems so that they evolve in new directions; Hayles invokes the imagery of a spring compressed to look like a circle that suddenly releases, and leaps into a new shape as symbol for third-wave cybernetics (Hayles 222). Like Heinz von Foerster served as a medial figure between first- and second-wave cybernetics, so too does Francisco Varela, Humberto Maturana's long-time associate, occupy a transitional position between second – and third-wave cybernetics. As Varela began working in a relatively new field, Artificial Life, whose approach to systems was to have the system's organisation not remaining static, but rather be able to "transform itself through emergent behavior" (Hayles 223). Echoing some of the disputes concerning the (dis)embodied nature of information in first-wave cybernetics, there are also diverse opinions in Artificial Life about the 'best' way to construct organisms; some researchers focus on simulations, arguing that embodiment is not necessary. Others, meanwhile, share the opinion that embodiment is a necessary component to "fully capture the richness of an organism's interactions with the environment" (Hayles 223) – a notion we shall return to when exploring "personhood". Maturana's objective observer which previously was a core idea, now retreats to the periphery, while having a role more akin to a narrator/narratee in stories concerning Artificial Life (Hayles 223). To further consider what Artificial Life is, let us proceed to evolutionary biologist, Thomas S. Ray.

In 1994 at the fourth conference on Artificial Life, Thomas S. Ray, submits two, at first glance offbeat, proposals:

1. The first one is a proposal with the aim of preserving biodiversity in Costa Rica's rainforests.
2. The second is a proposal to release his software program, Tierra, which is trying to create Artificial Life-forms on a computer, on the internet with the aim for it to spread to other computers thereby creating multiple 'species' on computers around the world. (Hayles 223-4)

For Ray these otherwise distant suggestions, are not so unrelated, in fact it is his view that the first proposal will help protein-based biodiversity in the world, and the second proposal will work towards the same goal of greater biodiversity – just in "silicon-based" lifeforms (Hayles 224). The intention of AL, as Artificial Life is commonly abbreviated as, is, according to Ray: "[…] to introduce the *natural* form and process of life into an *artificial* medium" (180; *emphasis added*). Ray's attitude to 'natural form' and 'artificial medium' reveals a belief that not only are computer codes alive but an even more enthralling idea: that they are natural. In other words: human creations – tools, really – are being used to create something natural and alive, through something that is hyperbolically "*a rock that we tricked into thinking"* (@daisyowl). Ray's assumptions are not outrageous within the field, but according to Hayles the answer to the question of whether or not they are alive and natural is a difficult one tied up with many assumptions "that are not explicitly articulated" (224). Instead of pursuing answers to these questions directly, Hayles proposes to find them indirectly through the stories told about them which she argues "constitute a multi-layered system of metaphoric and material relays through which 'life,' 'nature,' and the 'human' are being redefined" (224). She, specifically, identify three distinct narrative levels within AL:

1. The first level is explicitly concerned with Thomas S. Ray's Tierra program and other programs like it (Hayles 224)
2. The second level is fixated on the strategies and arguments employed by AL researchers to position AL as a "valid area of research within theoretical biology". However, according to Hayles, this level rapidly extends further than just "professional considerations", going on to contemplate life on Earth and potential kinds of life that might develop here (224).
3. Finally, the third narrative level concerns speculations on how human beings and the silicon-based 'creatures' located in computers will interact.

Before moving on to examining these three narrative levels in detail, Hayles elucidates on the three prevailing research spheres in AL, which the next paragraph will likewise do.

The three typical research fronts of Artificial Life are, according to Hayles, *wetware, hardware* and *software*. Below follows a short description of each term within the framework of Artificial Life, as all three are rather ubiquitous within many related fields.

- *Wetware* is characterised by being attempts to create artificial biological life, often through techniques employing "components of unicellular organisms in test tubes" (Hayles 225).

- *Hardware* is, in artificial life research, attempting to create and construct robots along with similar embodied life-forms (Hayles 225).
- *Software*, finally, is, according to Hayles, "the creation of computer programs instantiating emergent or evolutionary processes" (225). Software research typically approaches this by having "a few simple local rules" combined with highly recursive structures which creates environments where properties or programs can appear, often in unanticipated ways, unbidden (Hayles 225).

Not surprisingly the definitions of these terms here converge in their pursuit of creating a type of artificial life, but interestingly they each pursue and consider divergent avenues to achieve this goal. As Hayles formulates it: "they all share the sense of building life from the 'bottom up'" (225). Especially concerning software research, the emergence of spontaneous properties in the programs is not necessarily by themselves a type of living organism. Therefore, certain narratives that translate these emergences into evolutionary scenarios associated with behaviour typical to living organisms are unavoidable. Typically, these narratives function by taking the binary computer codes and transferring them in a way that makes sense but as biological counterparts, perhaps reminiscent of Darwinian evolution through survival, adaptation and mutation (Hayles 225). In this we can recognise Ray's proposal to release his Tierra program on the internet and subsequently computers connected to it.

Liberal Humanist Subject

Before diving fully into posthumanism, it is prudent to account for the 'humanism' part of posthumanism. The object of critique for posthumanism is the idea of the *liberal humanist subject* as the next paragraphs on posthumanism will also remark on. But what is the basis of this liberal humanist subject? Thomas Foster defines posthumanism in the following way:

> posthumanism defines itself relationally, as a redefinition of, or an alternative to, something construed as humanism. "Human" or "humanism" in practice names a set of sometimes overlapping, sometimes distinct objects of critique: variously, the liberal subject; romantic individualism; privatized psychic interiority ("thought" or "consciousness" considered as an exclusive property of human beings); human nature; essentialism; species-identity, understood as the basis for claims about anthropocentric uniqueness and superiority; or an overly idealized and abstract

> concept of "the" human body, as ground for species-identity and exclusionary definitions of the qualities that define the human. (Foster 452)

Posthumanism then identify itself as the opposition to a human-centric point of view that easily becomes a source of superiority. Steven Umbrello emphasises that the term 'humanity' "is a fragile concept – or even one that is illusory – built upon false notions of the necessity of human cognitive superiority" (29). Susan M. Squier likewise attributes postmodernism and the posthuman body to the opposition of traditional Western philosophy: "The dominant feature of postmodernism is its challenge to the master narratives of Western metaphysics and philosophy, with their bases in binary oppositions: mind/body; male/female; self/other; first world/third world; human/non-human" (119). Instead, posthumanism seeks to be less prescriptive than a more traditional prescriptive philosophical point of view would be. In the next few sections, it should become apparent how posthumanism can give space for more types of being to hold merit.

N. Katherine Hayles: How We Became Posthuman

Perhaps predictably, Hayles begins *How we became Posthuman* by discussing Alan Turing's Imitation Game test, commonly referred to simply as the Turing test. According to her the Turing test was to "set the agenda for artificial intelligence for the next three decades" (Hayles xi). A central tenant with the Turing test was the erasure of embodiment whereby "'intelligence' becomes a property of the formal manipulation of symbols rather than enaction in the human lifeworld" (Hayles xi). Embodiment is one focal point of Hayles' research, and it is likewise influential in posthumanism studies in general. We shall return to embodiment soon. With the erasure of embodiment, the importance was thought to be strictly on the "generation and manipulation of informational patterns" (Hayles xi). The subject of 'information' is also particularly central to cybernetics and subsequently posthumanism. Although both Hayles and other scholars have remarked on and have problems with Turing's first example in the Imitation game where the 'game' is not played between a human and a machine but rather a man and a woman, Hayles still sees some merit to the Turing test:

> Think of the Turing test as a magic trick. […] The important intervention comes not when you try to determine which is the man, the woman, or the machine. Rather, the important intervention comes much earlier, when the test puts you into a cybernetic circuit that splices your will, desire, and perception into a distributed cognitive system in which represented bodies are joined with enacted bodies through mutating and flexible machine interfaces. (Hayles xiv)

Through this interaction, Hayles argues, no matter what identifications you attribute to the embodied entities, human or otherwise, you are already posthuman (xiv). Posthumanism therefore is in its most basic form an examination of how humans interact with technology.

Embodiment

Hayles begins her examination of embodiment in the posthuman with what is to her a nightmarish thought: downloading human consciousness into a computer (Hayles 1), which she encounters in the notorious book by Hans Moravec, *Mind Children: the Future of Robot and Human Intelligence*. To Hayles the notion that human consciousness can be transferred into a container distinctly separate from the subject's original body and still remain utterly identical is shocking (Hayles 1). Pursuing this and other cultural phenomenon Hayles dives into, among other subjects, information theory, research on virtual reality, computer simulation, cybernetics and cognitive science. Through this submersion she points to three "interrelated stories" in the aforementioned subjects (Hayles 2). The three stories are summarized here:

1. "Information lost its body". In other words, information is considered separate from its material medium in/on which it was embedded.
2. "The cyborg was created as a technological artifact and cultural icon".
3. "The human is giving way to a different construction called the posthuman". This is deeply engrained with the previous two stories and a still unfolding story according to Hayles (2).

Already with some of the first cybernetic scholars the idea of information being separate from the liberal humanist subject that Hayles concerns herself with became a widespread view. Commonly the theme of the posthuman became the union between human and the intelligent machine (Hayles 2)

Moving on, Hayles naturally asks "what is the posthuman"? She suggests a list of four elements that are prevalent assumptions in the view of what comprises the posthuman:

1. The posthumanism view lays more emphasis on informational patterns as opposed to the material. Through this view, embodiment is also seen as "an accident of history rather than an inevitability of life" (Hayles 2).
2. Consciousness, regarded as the seat of human identity in Western tradition, is considered "as an epiphenomenon, as an evolutionary upstart trying to claim that it is the whole show", while it is actually just a "minor sideshow" (Hayles 2).

3. The body is considered "the original prosthesis" that we have been learning to use since the moment we are born. Therefore, replacing/developing on the body is just continuing an ongoing process with the body.
4. Finally, these and other views shape the idea of a human being that can be joined with intelligent machines with harmony. In essence, the posthuman view does not recognise any distinct boundaries between "bodily existence and computer simulation, cybernetic mechanism and biological organism, robot technology and human goals" (Hayles 3)

Although, the notion of the body as a prosthesis and the seamless joining of the human being with intelligent machines effortlessly conjure examples of cyborgs, such as Robocop, it is not a prerequisite that the posthuman subject is a literal cyborg. Hayles underlines that fields like cognitive science and artificial life considers the construction of subjectivity the defining characteristic of the posthuman subject, thereby rendering the natural, biological body just as posthuman as that of a cyborg (4).

Continuing with the liberal humanist subject and embodiment, cybernetics has some similarities in perspectives with the criticisms raised by a number of diverse perspectives: feminist scholars pointing out that the liberal humanist subject has historically been considered a white, European male; postcolonial theorists likewise take issue with the historically white male liberal subject along with the notion of a universal, consistent identity. Postmodern scholars like Gilles Deleuze and Felix Guattari linked the liberal humanist subject with capitalism. Cybernetics, however, chiefly concerned itself with explaining the human being as informational processes (Hayles 4). As information has lost its body, this view indicates that the embodiment is not fundamental for the 'construction' of a human being. In fact, the erasure of embodiment as an essential component of the liberal humanist subject has not occurred in either feminist or postcolonial theories (Hayles 4). Quoting William Gibson's *Neuromancer*, Hayles explains the posthuman subject as: "Data made flesh" (Gibson 8) in that posthumanism constructs the "body" through information/thought (Hayles 5). However, she also emphasises her dreams for posthumanism:

> […] that embraces the possibilities of information technologies without being seduced by fantasies of unlimited power and disembodied immortality, that recognizes and celebrates finitude as a condition of human being, and that understands human life is embedded in a material world of great complexity, one on which we depend for our continued survival. (5)

In these next few paragraphs, the focus will be exclusively on the body and embodiment. The notion that embodiment is unneeded is not an uncommon ideology in cultural theory, according to Hayles (192). Hayles suggests "a new, more flexible framework in which to think about embodiment in an age of virtuality", which is made up of two interacting polarities (193).

1. The first polarity is constructed through an interplay between the cultural constructed body and the experiences that each individual "within a culture feel and articulate" (Hayles 193).
2. The second polarity Hayles understands as a polarity that emerges from inscribing and incorporating practices (193).

To begin to explain her understanding of the term *embodiment*, Hayles echoes the scholar Elisabeth Grosz, when she states that "there is no body as such; there are only *bodies*" (qtd. in Hayles 196; emphasis in original). Embodiment contrasts the concept of the body in that the body is understood to be normative within a predetermined group of criteria (Hayles 196). In other words, the body is a collection of specific data that then produces a useful concept that is a body. Contrasting this, is embodiment which is always dependant on context: "enmeshed within the specifics of place, time, physiology, and culture, which together compose enactment (Hayles 196). Judith Halberstam and Ira Livingston further elaborate on their notion of the posthuman body: "Posthuman bodies were never in the womb. Bodies are determined and operated by systems whose reproduction is - sometimes partially but always irreducibly - asexual: capitalism, culture, professions, and institutions, and in fact sexuality itself" (17). Therefore, while 'the body' and embodiment overlap each other, they will never correspond completely. By its nature it is always difficult to completely make an account for a specific embodiment as embodiment is as varied as the people who exist.

Moving on then to elaborate on her second polarity, inscribing and incorporating practices will be examined. Hayles uses these terms as developed by Paul Connerton in *How Societies Remember*. *Inscription* or an *inscribing* practice is similar in nature to the concept of 'the body' in that it is "normalized and abstract" (Hayles 198). An *incorporating practice* is explained by Hayles as "an action that is encoded into bodily memory by repeated performances until it becomes habitual" (199). One example of an incorporating practice is learning to type. Claiming that someone can type means that the person is able to account for how the keys work or indeed where they are located specifically. Indeed, the shared understanding behind this statement is that the person has performed the action of typing until the keys are as an extension of their fingers. Interestingly, this person can type something

without being able to read it, like for example a different language. Incorporating practices can be summarized as "always performative and instantiated" which, as Hayles notes, also means that they are always context specific (200). Since no embodied individual is completely alike, the exact manifestation of the embodied individual impacts how the incorporating practices are expressed. Incorporating practices are therefore by necessity, as opposed to inscription, always somewhat dependent on context. Another important observation Hayles makes, is that incorporating practices "are [not] in any sense more 'natural', more universal, or less expressive of culture than inscribing practices" (200). The 'body' learns culture, becomes *enculturated*, through both practices. Bodily practices such as sitting, walking, gesturing and indeed moving are culturally specific and largely learned. Certainly, expression such as "girls don't sit with their legs open" and "boys don't walk like that" demonstrate how the body's incorporating practices are further encultured by inscribing practices in gendered performance within a certain culture where these practices are significant (Hayles 200). Following this examination of body/embodiment and inscribing/incorporating practices, the next section concerns itself with the theoretical framework for analysis literature that Hayles' attempts to establish.

Consciousness and Posthumanism

The next paragraphs will dive further into Hayles a proposed theoretical framework. Her exploration of posthumanism and consciousness with a quote from Ihab Hassan predicting posthumanism entrance to human consciousness:

> We need first to understand that the human form – including human desire and all its external representations – may be changing radically, thus must be re-visioned. We need to understand that five hundred years of humanism may be coming to an end as humanism transforms itself into something we must helplessly call posthumanism. (Hassan 843)

Using the cybernetic terms that Hayles has explored so far, she now attempts to use them to "map the posthuman as a literary phenomenon" (Hayles 247). As she points out one of the issues concerning posthumanism is that, although it is a fairly recent concept, it already involves an extensive number of cultural and/or technological spheres, "including nanotechnology, microbiology, virtual reality, artificial life, neurophysiology, artificial intelligence, and cognitive science, among others", which makes mapping out posthumanism exactly challenging (Hayles 247). Yet, despite this, Hayles attempts to construct a map of the posthuman, starting with the, according to her, two central dialectics of posthumanism: presence/absence and pattern/randomness (Hayles 247), which we shall now turn to.

Hayles remarks that as information becomes more significant and important, the pattern/randomness dialectic, which is deeply tied to information, becomes predominant over the presence/absence dialectic (Hayles 247). While this might be the case, she argues, it does not eliminate the need for the explanatory power of the dialectic of presence/absence as this "connects materiality and signification in ways not possible within the pattern/randomness dialectic. To be useful, the map of the posthuman needs to contain both dialectics" (Hayles 247-248). It is, therefore, advantageous to view these two dialectics as mutually beneficial regarding the theoretical framework that Hayles proposes. She begins the process of describing these in detail by imagining the presence/absence – and pattern/randomness dialectics as two axes of a semiotic square. Using a semiotic square appeals to Hayles, as its combination of structure and flexibility allows for an easy overview of how the dialectics interact within themselves but also with each other, finally also allowing for synthetic terms to emerge as well. In practice, Hayles places the presence/absence dialectic on the primary horizontal axis, with the pattern/randomness dialectic being on the secondary horizontal axis below it. The relation between the two dialectics is characterised by "one of exclusion rather than opposition", according to Hayles (248). They each tell a story that the other cannot know the same way. Furthermore, Hayles connects the two dialectics with each other diagonally and labels the presence/pattern connection *replication,* as these two notions share replication as a theme: "an entity that is present continues to be so; a pattern repeating itself across time and space continues to replicate itself" (Hayles 248). Similarly, the absence/randomness notions likewise share a theme, *disruption*: "absence disrupts the illusion of presence, revealing its lack of originary plenitude. Randomness tears holes in pattern, allowing the white noise of the background to pour through" (Hayles 248). Hayles' original square on *the semiotics of virtuality* (figure 2) can be seen below where all these connections are plotted in.

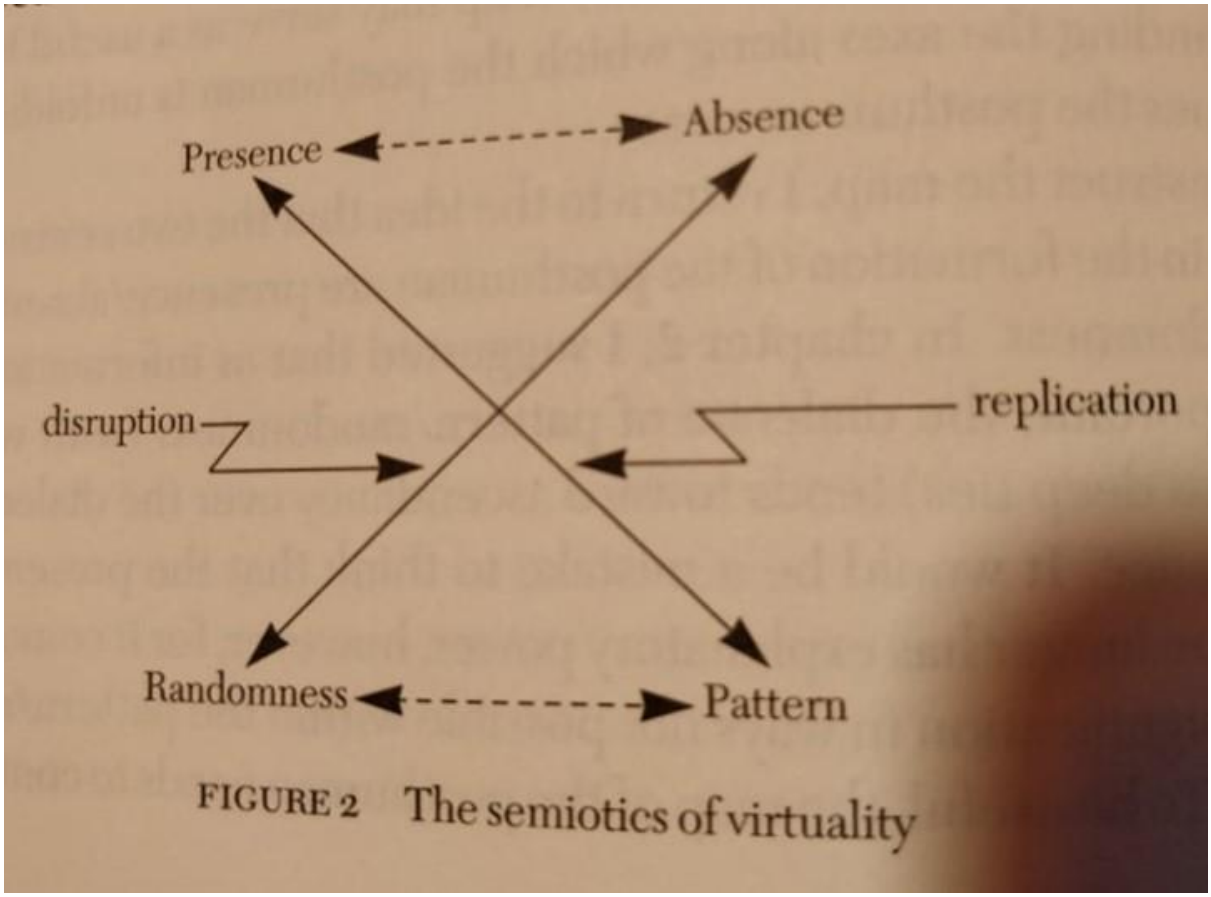


FIGURE 2 The semiotics of virtuality

As Hayles states previously, one of the advantages of the semiotic square is the flexibility of it, and that is what she now employs by putting it dynamically in motion, because the interplay between the primary- (presence/absence) and secondary axes (randomness/pattern) further produce other dialectics, which then again can produce even more dialectics ad infinitum. Still, Hayles only includes the next 'layer' after her original square in her examination of the posthuman. The 'new', transformed semiotic square can be seen below, but the added layer will be gone over first. All the terms the emerge from the intersection of the original dialectics, are what Hayles call "synthetic terms". They are as follow:

- Top horizontal interplay between, presence-absence, produces *materiality.*
- Left vertical interplay between, presence-randomness, produces *mutation.*
- Right vertical interplay between, absence-pattern, produces *hyperreality.*
- Finally, bottom horizontal interplay between, randomness-pattern, produces *information.*

Although, these synthetic terms are quite unequivocal they will be elaborated on in greater detail. *Materiality* Hayles defines as: "I mean the term to refer both to the signifying power of materialities and to the materiality of signifying processes" (249). *Mutation* comes about through the specific interaction between randomness and presence since that is the mark randomness leaves on presence; when a random event happens to a piece of DNA, it will potentially change the manifestation of the DNA in the world itself. With the term *Hyperreality,* Hayles largely echoes Jean Baudrillard's hyperreality, the process which he, according to Hayles, described as: "a collapse of the distance between signifier and signified, or between an 'original' object and its simulacra" (Hayles 249). This idea of simulacra is not necessarily a simple concept to grasp, but it can be simplified as a process where the prevalence and visibility of reproductions of a 'thing' displace the original. The example used by Hayles is that of the Mona Lisa and a lifetime of encountering reproductions, leading to *the* original painting being seen as simply another reproduction instead of the original (Hayles 250; emphasis in original). Finally, *information* is intended by Hayles to mean "both the technical meaning of information and the more general perception that information is a code carried by physical markers but also extractable from them" (250). To flesh out her

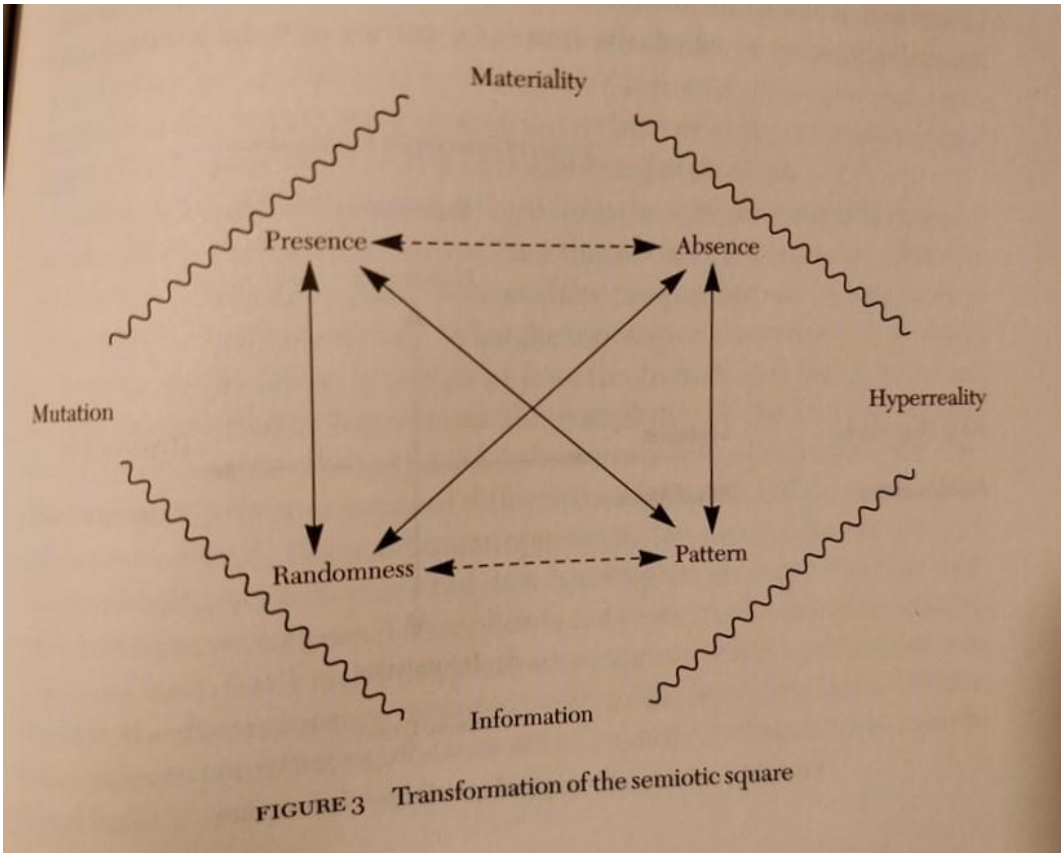


FIGURE 3 Transformation of the semiotic square

transformed semiotic square, Hayles introduces four tutor texts: *Galatea 2.2., Snow Crash,*

Blood Music and *Terminal Games*, which will not be directly discussed here, but her overall interpretation of the posthumanist subject, will be the focus of the next paragraph.

Through her examination of the aforementioned four texts, one recurrent trait of the liberal humanist subject manifests itself as a posthuman attribute across all four texts, namely that of *agency* (Hayles 279). According to Hayles, "the posthuman tends to be embraced if it is seen as preserving agency (*Blood Music*) and resisted if not (*Terminal Games*)" (279). Additionally, the texts reveal that the construction of the posthuman is thoroughly interested with the boundaries that challenge and change the notion of selfhood. The constructions of the posthuman in Hayles four texts all lead her to claim that consciousness is fragile. In can be high-jacked, erased, absorbed and back-propagated through flawed memory. As consciousness is increasingly seen as being comprising multiple coding levels, it becomes increasingly vulnerable to possible sites of intervention that can have disastrous consequences (Hayles 279). What this means for the posthuman subject, is that one can no longer assume that consciousness ensures the existence of a "self", which Hayles argues also means that the posthuman subject is also a "post-conscious" subject" (280).

Summarising her findings on the posthuman in the four tutor texts, Hayles posits the following assertations: they are "obsessed" with the question of evolution/devolution. Even more profound though is the underlying question all the texts express:

> When the human meets the posthuman, will the encounter be for better or worse? Will the posthuman preserve what we continue to value in the liberal subject? Will free will and individual agency still be possible in a posthuman future? Will we be able to recognize ourselves after the change? Will there still be a self to recognize and be recognized? (Hayles 281)

Although each of the texts ask themselves variations of these questions, they also, surprisingly, all seem to be continuously committed to "some version of the human subject" (Hayles 281). These questions are remarkably close to the questions posed by this thesis, and for some of them a possible answer emerge through the discussion

How We Became Posthuman – Concluding Findings

Hayles initiated *How We Became Posthuman* by propounding that "the prospect of becoming posthuman both evokes terror and excites pleasure" and now at the end of her exploration of the posthuman, she returns to expound on this terror/pleasure duality within the posthuman. Beginning with the easier of the two, Hayles argues that the terror aspect associated with the posthuman is relatively easy to understand (283); the *post-* suffix making up part of 'posthuman' produces an implication that the "days of 'the human' may be

numbered" (283) with some scholars believing this in a quite literal sense of humanity as the dominant life on Earth being displaced by intelligent machines, with Moravec's *Mind Children* being a major inspiration. With this outlook producing the, admittedly, poor choices of either accepting that humanity's fate is to fade away like so many species before us, or otherwise extend our time on the planet by joining the intelligent machines, by becoming machines ourselves. As for pleasure, the posthuman represents an opportunity for Hayles and likeminded individuals to analyse and think about what it means to be human in less conventional ways. One of those ways is Anne Foerst's discussion of a less restricted notion of *personhood*, which the following section will examine.

Anne Foerst: Understanding Personhood

While the necessity of the embodied human in posthumanism has been covered extensively, the obvious questions someone might ask themselves now is: why does the sentient machines' physical appearance often seek to imitate human bodies? In this next section, this question will be considered. In her article "Artificial sociability: from embodied AI toward new understandings of personhood" Anne Foerst deliberately avoids treating technology as "Other", while accounting for why the artificial creatures being developed at the MIT Artificial Intelligence Laboratory are specifically created to look humanoid (373). In her words "the focus of attention [on the evaluation of AI technology] should be not on some presumed 'otherness' but on the underlying presuppositions about the functionality of the human biological system" (Foerst 373). Which she elaborates on as moving away from discussing the dangers of AI, whether through loss of jobs or dehumanisation in the classroom when students interact with learning machines, but "toward an ethical evaluation of the underlying anthropology of AI which already shapes human self-understanding in the Western world" (Foerst 373-4).

Specifically, in her accounting of the situation with humanoid robot projects pursued around the globe, Foerst highlights how (at the time of writing) this research was predominately carried out in Japan with the intended goal of alleviating the economic burden of a population with a rising old age dependency ratio. Therefore, in this instance the robots are being developed with as humanoid of an appearance as possible for two reasons:

- First, according to Foerst, the people that the robots work for will have an easier time accepting the robots if they appear as human-looking as possible. They will be able "to project into it [the robot] phenomena such as friendship, warmth, empathy, etc." much more readily (374).

- The second reasoning, according to Foerst, is of a more practical nature. If the robots are going to be working in houses designed and built for humans, it is naturally desirable if the robots share humanoid features that would allow them to more easily navigate human homes (374).

These two explanations for robots having human-like appearances obviously support each other, but particularly the first argument: that humans will easier be able to project human qualities onto the robots if they appear like humans is particularly interesting. Practically, it makes sense to design robots that suit their environments, which the two texts will later show some considerations of. The projection of human qualities onto synthetic life, robots and the like is, likewise, a major point in both *Consider Phlebas* and *Software*, and as such it is something this paper will return to discuss.

Turning then to the MIT Artificial Intelligence Lab, where researchers are trying to construct two "babies" called *Cog* and *Kismet* under a certain AI camp with "a new set of assumptions about the nature of intelligence", with the following four headings: embodiment, interaction, development and integrated architectures (Foerst 377). They can be summarised in the following way:

- Embodiment: Both AI projects begin with the insight "that it is impossible to abstract intelligence from bodily features and conditions. Intelligence, according to the embodiment thesis, cannot be implemented on a disembodied machine because it emerges only in minds that are embedded in a world" (Foerst 377). Taken further, this means that specifically a human-like intelligence can only come about in a body that is as human-like as possible, since it will live and experience the world like humans which also feeds into the next assumption.
- Interaction: The functions, forms and signals of the body are essential for the formation of intelligence. The body is the original tool of the intelligence, which enables the intelligence to engage in the real world; to understand the signals it receives and to construct relationships with other intelligences. For this reason, it was also deemed necessary to construct the Cog and Kismet machines with a wide-ranging array of sensors.
- Development: Following the logic of the researchers calling their AIs "babies", it is no surprise that the assumption here relates to intelligence not emerging "ready-made" (Foerst 377). As the AIs interact with the world through their bodies, they develop the skills to do more complex tasks through the repetition of simple tasks.

- Integration: Finally, to facilitate the first three assumptions, every part of the AIs that the researchers are developing is a small system that can be considered both independent and connected to all the other systems at the same time. As these systems learn to interact with each other, the possibility of higher-level behaviour emerges.

Interestingly, what emerges from the construction of Kismet and Cog is not just practical experience, but also a discussion of the definition of "personhood", which we shall now briefly turn to.

Foerst begins this topic by accounting for different approaches to the question of *cognition* in cognitive science. The classical view within AI approaches is that *cognition=software*, "which can either run on the *hardware* of the computer or the *wetware* of the brain, while being independent of both. Cognition is thus separated from bodily realities, from emotions and other features that do not fit this metaphor" (Foerst 379). This metaphor of the body/cognition relations is widely influential. Positioned against this *cognition=software* approach, we find, according to Foerst, a paradigm shift in cognitive science in the form of embodied AI: "emotions now are recognized as pre-requisites for cognition, the importance of the individual bodily reality is accepted, and several brain studies demonstrate the correlation of brain architectures and neural processes on the one hand, and cognitive abilities and features like self-consciousness on the other" (380). One development following this changing approach to cognition is the impact on "our intuitive self-understanding", as this new approach reduces the differences between constructed AI and organic humans. Before there is clear distinction between the wetware of the brain and the hardware of the computer enabling the function of software. However, by viewing cognition in relation to emotions these differences are less important. With the differences between intelligent machines and robots seemingly becoming minuscule, Foerst considers it prudent to "ask for criteria to define what it means to be a person" (381), and we shall now try to account for these.

Foerst indicates several distinct possible criteria for defining "personhood": *genes*; as an entity separated from their *environment* and finally utilising the concept of "*character*". Genes, although they are unique to every person (excepting identical twins), do not adequately account for what makes up a person, states Foerst, as a large part of personhood is also defined by culture and our upbringing (381). Likewise, defining personhood as caused by environment, is rejected by Foerst: "an entity separated from her environment. A person, one might argue, is distinctly herself and the boundaries between self and others and/or environment is unambiguous" (381), due to the advent of technology that changes a person's

interaction with their environment. In this case: glasses, plastic surgery or organ transplantation are all examples of technological steps that can change the personality of a person (381), to different extents. Generally, technology labelled by Foerst as "cyborg technology" has a profound ability to impact the personality of a human (382). Finally, Foerst proposes to use the idea of "character" to define what makes up a person – what makes them an individual. Unfortunately, as she points out, there is currently no way to tell what part of our genetics or development that influences which part of our "character", so this approach is likewise moot, rendering all three approaches to define a person empirically futile (382).

With AI growing ever closer to human representation, the possibility of highly human-like AI also enters the discussion of a potential understanding of personhood. This leads Foerst to conclude that there has been plenty of examples throughout history of (human) people being excluded from social intercourse through various cultural or physical characteristics, therefore she would "much prefer to adopt a very broad, non-empirical concept of personhood, which might include robots and some apes, rather than run the risk of excluding some humans from the community" (385). Approaching personhood with a broad definition in mind, as Foerst proposes, gives the impression of a sensible attitude to a, if it is not here already, fast-approaching posthuman world, where the lines between human, posthuman and artificial life become progressively blurrier. This struggle for identification as an individual that exists, a person, is a conflict reflected in both *Consider Phlebas* and *Software*, although they approach this theme in divergent manners. Lastly, we will now turn to provide some context into transhumanism.

Hans Moravec – 'Mind Children' and Transhumanism

As previously observed Hans Moravec is a name that instantly invokes nightmarish imagery for posthumanists like N. Katherine Hayles and other scholars. Although, Moravec's influential book *Mind Children: the Future of Robot and Human Intelligence* is generally considered as belonging to the area of transhumanism, it seems appropriate to devote a little effort to elaborate on Moravec to illustrate where posthumanism and transhumanism ideas diverge for they obviously harmonise to some extent. If nothing else the naming of posthumanism and transhumanism does not make the somewhat divergent ideas opaque.

Moravec lays his view of the future of humankind clear immediately in the prologue: "What awaits is not oblivion but rather a future which, from our present vantage point, is best described by the words 'postbiological' or even 'supernatural'. It is a world in which the human race has been swept away by the tide of cultural change, usurped by its own artificial progeny" (1). The future, according to Moravec, is one in which humanity has successfully

created machines that are surpassing ourselves, eventually "transcending everything we know" (1). The relationship between us, he likens to that of parenthood and children, where our children will seek their own destiny while humanity, "their aged parents, silently fade away" (1). Therefore, it is hardly surprising that Moravec considers how humanity might keep up with our artificial offspring. Although, genetic engineering might hold one answer to improve the human body, from intelligence, to metabolism to extending the natural lifespan of the body, the medium of a protein-based body is in Moravec's view unsatisfactory to keep up with our artificial children (108). Instead, he envisions a process whereby our mind could possibly be transcribed out of the brain to be turned into a sort of computer program: the concept of *Transmigration*.

This is the 'nightmarish' imagery Hayles describes earlier. The merit of the transmigration stems from Moravec's distinction between the brain and the human mind. The brain serves the function of the hardware, or in this case more accurately the wetware, on which the software, the mind, functions. Thus, if the mind can be scanned and transferred to a computer, your biological brain will eventually die, perhaps even as the transfer is completed, while your mind will live on, separate from your original human body. In Moravec's view, it will then be possible to enhance the mind further than what the biological human body is capable of (112). With the flexibility of this new mind-program, you will be able to download copies of your mind into whatever body is suitable to experience environment you wish. Should you wish to, it will even be possible to have multiple copies of your mind walking around in different bodies at the same time. Thus, Moravec concludes that the permanent death of any one person is highly unlikely (114).

For Moravec downloading the consciousness, is the primary tool for future proofing humanity, while Hayles and other posthumanists see the consciousness and body as inseparable halves making up the whole. In fact, the human body is seen as the "original prosthesis", shaping the individual. Here, we see the crux of Hayles' dread concerning Moravec and his certainty in a biologic humanity's extinction. Though, Moravec acknowledges that the mind will take time to adjust to its new circumstances in a robot body many times more efficient than previously, it seems a minor concern to him (112), whereas the posthuman view is much more dubious of such a claim as we have seen.

While Moravec's ideas about the future of humankind presented so far are somewhat disconcerting, he does also make a case for a more posthumanist viewpoint in the chapter aptly titled *Symbiosis*. For at least the foreseeable future Moravec images the human/robot interaction as less transhuman and recognisably posthuman: As the machinery grows in

flexibility and initiative, this association between humans and machines will be more properly described as a partnership. In time, the relationship will become much more intimate, a symbiosis where the boundary between the "natural" and the "artificial" partner is no longer evident" (75). In this chapter he envisions scenarios, where the marriage between humanity and the body are articulated in a symbiosis. Some of it the scenarios dreamt up by Moravec, involve scenes that will be instantly recognisable to anyone with even a passing familiarity with VR, *Virtual Reality*.

Rubin – refuting Moravec and Transhumanism

In the course of reading N. Katherine Hayles, the influencing thoughts of Hans Moravec's view on the future of humanity seem particularly pervasive within posthumanism (and decidedly transhumanism). Therefore, the previous section devoted attention to Moravec and his book *Mind Children: The Future of Robot and Human Intelligence*. With the following paragraphs the aim is to briefly consider some criticisms on Moravec's position on the future of humanity and intelligent machines. Apart from Hayles' already discussed opposition to Moravec, this section turns to Charles T. Rubin's article "Artificial Intelligence and Human Nature".

In it, Rubin starts out his discussion by boldly claiming that: "The cutting edge of modern science and technology has moved, in its aim, beyond the relief of man's estate to *the elimination of human beings*" (Rubin 88; emphasis added). This position, Rubin claims, is not a new development, which is not what he personally takes issue with. His opposition to this claim stems from the fact that: "[…] the proposals of a small, serious, and accomplished group of toilers in the fields of artificial intelligence and robotic. Their goal simply put, is a new age of post-biological life, a world of intelligence without bodies, immortal identity without the limitations of disease, death, and unfulfilled desire. Most remarkable is not their prediction that the end of humanity is coming but their wholehearted advocacy of that result" (Rubin 88). Later, he goes on to discuss this group of thinkers collectively as "extinctionists" (Rubin 90). Accordingly, humanity's extinction will be accomplished through a dual combination of humanity choosing to voluntarily to undergo a transformation into machines while simultaneously "losing out in the evolutionary competition with machines" who are already faster than the biological human brain (Rubin 89). Rubin finds nearly all these elements problematic as they tend to involve primarily metaphysical-, technical- or philosophical speculation that are highly debateable (89). One of the driving forces behind this belief is, according to Rubin, Hans Moravec and his *Mind Children: The Future of Robot and Human Intelligence*, which he labels as "perhaps the ur-text of 'transhumanism,' the

movement of those who actively seek our technology-driven evolution beyond humanity (91). If we take this view of "transcending" normal mortality as definition of transhumanism, the line between posthumanism and transhumanism is simultaneously distinct and blurred. As this thesis focuses on posthumanism, transhumanism will largely be ignored from this point on, other than the previous accounting for Moravec's nightmarish/enthusiastic foretelling of human extinction/ascension.

Generally, Rubin highlights several issues concerning humanity's supposed ascension to robots, but there is one issue in particular that is instantly recognisable for those familiar with Hayles and posthumanism: the *unproblematic* loss of the human body. Rubin takes issue with how Moravec seemingly treats bodies "as a trivial component of personality; after all, they change dramatically over time and we do not lose our sense of identity as a result" (94). However, this argument is flawed in Rubin's view. He points to the fact that a 16-year old 'I' is not the same as a 45-year old 'I' even should the 45-year old 'I' claim to "still feel 16 inside" (Rubin 94). Not only is the difference between the two ages obviously going to be reflected simply in physical differences, but they also "involve a deeper transformation of our longings, our understanding of the world, and our duties that cannot be separated from our existence as embodied creatures" (Rubin 95). Even should the aforementioned 'I' live on in a replacement robotic body, the changes between the sensory input and the mechanisms processing them along with the ability to make use of the virtual world leads to an insurmountable chasm between the experiences and point of view between the biological 'I' which started out as the embodied being and the new being governed by its new software and hardware. To Rubin the closest analogy to the likeness of the relationship between the machine and the original embodied being is the relationship between adults and infants (95). The incredible difference of perspective leads Rubin to conclude that the robot world envisioned by the extinctionists will not be a human world. It is inconceivable that the human being of today would appreciate living in the machine-friendly world envisioned by the extinctionists (95). Precisely the topic of the robot world, is something quite obvious in both *Software* and *Consider Phlebas* where the reader encounters worlds specifically used by synthetic life.

***Consider Phlebas,* Analysis**

Having considered accounted for the posthumanist framework, this part of the thesis will turn its attention to the two texts of singular interest: Iain M. Banks' *Consider Phlebas* and Rudy Rucker's *Software*. Before delving further into the analysis, there will be some preliminary remarks on the structure of the following analysis sections. *Consider Phlebas* and

Software will be considered and quoted separately from each other in these sections of the paper. I will start with Banks' *Consider Phlebas* and then move on to treat Rucker's *Software.* To introduce each text, both works will be briefly summarised, and any interesting or unusual facets of each work highlighted. Lastly, both texts are part of a larger body of work with Banks' *Consider Phlebas* being the first book set in his *Culture*-universe, which spans full-length 10 novels. Similarly, Rucker's *Software* is also not a standalone novel, but the first novel in the larger *Ware Tetralogy*. With those initial considerations out of the way, I will now turn to *Consider Phlebas.*

Consider Phlebas

Consider Phlebas is Iain M. Banks' first science-fiction novel, published in 1987. It is part of the larger *Culture*-series set in the same fictional universe. On Banks' own webpage, the novel is described as "a space opera of stunning power and awesome imagination, from a modern master of science fiction" (Iain-Banks.net). The story of *Consider Phlebas* takes place in the middle of a galactic war between the zealous Idirans and the diverse Culture. The story predominantly follows the Changer, Horza, who fights on the side of the Idirans, as he is sent off to capture a newly constructed Culture *Mind*, a sentient supercomputer, for the Idiran war effort.

Of immense interest is the essay written by Banks himself "A Few Notes on the Culture" that was posted to newsgroup rec.arts.sf.written in August 1994. In it, Banks elucidates and provides context on varies different topics related to his Culture-universe. Among other things, he discusses various reasonings that lead him to envision the Culture as a post-scarcity society, developed through a planar economy. Related to this, Banks also discusses the unique circumstances that a society capable of living and traveling in space face; specifically, the independent nature of spaceships and habitats in space are at odds with traditional hegemonies' control mechanisms (Banks "Notes"). However, while these and other topics raised in this essay are tremendously fascinating, the thoughts of most consequence to this paper are surely Banks' thoughts in these 'notes' on the AI/human relationship in his Culture universe. Banks is decidedly positive about of our chances of eventually creating sentient AIs: "[AI] is taken for granted in the Culture stories, and […] is not only likely in the future of our own species, but probably inevitable (always assuming homo sapiens avoids destruction)" (Banks "Notes"). He, likewise, remains positive that AI would choose to work with "their source civilisation", and furthermore, he makes an eminently accurate observation in relation to this thesis: "At this point, regardless of whatever alterations humanity might impose on itself through genetic manipulation, humanity would

no longer be a one-sentience-type species. The future of our species would affect, be affected by and coexist with the future of the AI life-forms we create" (Banks "Notes"). While "A Few Notes On The Culture", has several more useful observations, explanations and hypotheticals, I will now turn to *Consider Phlebas* and outline some of the specific areas of interest in *Consider Phlebas* in conjunction with this thesis:

- The Changer, Bora Horza Gobuchul, who can consciously alter his biological body to suit a vast variety of situations.
- The Culture and their sentient Minds.
- Life and sentient machines.

The next few sections will be separated according to the above with each topic being discussed independently.

The Changer

Horza is the main character of *Consider Phlebas,* and, as will become apparent, he has an unmistakeable bias against the Culture, especially against sentient machines, which we will return to in the section 'Life and Sentient Machines.

We will now turn to consider some aspects of Horza himself as his existence and that of *Changers* in general is somewhat unique to the Culture universe.

The Changers themselves, while not explicitly aligned with either the Culture or the Idirans, are from an asteroid that technically lies within Idiran-controlled space. They are a race of humanoid-like people with an extreme degree of conscious control over their bodies. This makes them excellent at infiltrating other societies and general espionage. The aim of the next paragraphs is to detail and examine the capabilities of the Changers and their view on their own bodies.

We first encounter the protagonist of *Consider Phlebas,* Bora Horza Gobuchul, on the planet Sorpen. The reader's first impression of Horza and his somewhat unique body is conveniently told to us through his imprisonment and subsequently expected execution by drowning in literal waste in the dungeons. In this scene, Horza remarks that, although, he has turned off his ability to feel pain in his wrists and shoulders, there is still the nagging feeling in the back of his skull that he "*ought* to be hurting" (Banks *Phlebas* 11; emphasis in original). At the same time, he slips into what he calls the "Changing trance", which we learn allows him to make these incredible changes to his body (Banks *Phlebas* 14). Something that is seemingly possible for all changers, but not other humans/humanoids. The reader also learns that Horza is in this predicament because he has murdered and assumed the physical

appearance of a so-called "outworld minister". The mimicry is so convincing that the official coming to gloat over his imminent demise cannot tell him apart from the original person/body until he starts talking, after which his manner of speaking to his captors makes it obvious that he is someone else (Banks *Phlebas* 12). Along with changing his body to that of a native Gerontocrat, he additionally seems to have given himself highly venomous nails and teeth in preparation for his covert infiltration of the Gerontocracy of Sorpen. Due to the Changers' ability to convincingly shapeshift and assume another's appearance, Horza's captor labels the Changers "dangerous frightening things!" (Banks *Phlebas* 12).

The sentiment of hostility towards Changers expressed by Horza's captor on Sorpen is by no means an isolated notion, which this paragraph will explore in detail. Upon his chance arrival onboard the *Clear Air Turbulence*, Horza is made to kill for his spot on the ship. It is during this altercation where Horza deliberates on his tools for survival;, Horza still has his venomous nails, but while they would surely let him kill his opponent, the signs of venom would easily give him away as a Changer. He considers this certain death, as he reflects:

> A Changer was a threat to anybody who ruled by force, either of will or of arms. […] There was also a degree of human-basic revulsion reserved for Horza's species. Not only were they much altered from their original genetic stock, they were a threat to identity, a challenge to the individualism even of those they were never likely to impersonate. It had nothing to do with souls or physical or spiritual possession; it was […] the behaviouristic copying of another which revolted. Individuality […] was somehow cheapened by the ease with which a Changer could ignore it as a limitation or use it as a disguise. (Banks *Phlebas* 46-47)

There are two fundamental principles that Horza expound in the above paragraph. First, that people in power fear the Changer's ability to impersonate them. Second, that humans abhor the threat to their basic belief in their own identity that apparently having an ostensibly unique body gives them. The Changer's impersonation of the individual is thus a violation, perhaps *the* violation, of the individual. Coincidentally, both of the people, which Horza impersonate in *Consider Phlebas* the Gerontocrat, Amahain-Frolk, and the captain of the CAT, Kraiklyn, who are ruling through force, end up casualties of Horza's plots to impersonate power and use it, so their fear might not be entirely unfounded. Likewise, while it certainly would have disguised him better if he had tried to impersonate someone else onboard the CAT, Horza chooses to go back to an appearance that closely resembles his 'neutral' state rather than impersonate someone else. Further evidence supporting Horza's assertation that other humans loathe Changers as a people is the Changer people's fate, as it is

revealed just before the epilogue of the novel, in the 'Dramatis Personae' section of *Consider Phlebas*: "The Changers were wiped out as a species during the final stages of the war in space" (Banks *Phlebas* 467). With the fate of the Changers unveiled, the thesis will now turn to the aforementioned 'neutral state' and other facets of Changer culture.

As the readers quickly learn during Horza's imprisonment and later discussions with the Balveda, the Changers are ostensibly a unique type of human in the Culture universe with their ability to manipulate their bodies to the extent that they can. Unsurprisingly, their talents and the consecutive mistrust from their fellow humans makes for an interesting culture. Although, we do not learn much of the Changers other than that which we see through Horza's eyes, there are some allusions to the characteristics of this people. Firstly, there is the notion of a "normal, neutral state" as Horza calls it (Banks *Phlebas* 20). This neutral normal is the natural expression of the Changer's body: "In a Changer's mind there was a self-image constantly held and reviewed on a semi-subconscious level, keeping the body in the appearance willed" (Banks *Phlebas* 20). Whatever shape other than his/her natural expression is one that a Changer has chosen. But the change is not necessarily just skin-deep, as we learn when Horza contemplates his options in assuming Kraiklyn's appearance. Here he reveals that Changers can even copy the DNA of the people that they impersonate with the only organs unaffected being the brain and optionally the gonads (Banks *Phlebas* 105). It seems then that the instruments for procreation, to a lesser extent, and the house of the mind are protected from Changing. While the mind and genetic legacy of the Changer undergoing modifications thus both seem secure, there are additional complexities to the Changers' own culture.

As he demonstrates on the CAT, Horza has no qualms about killing other human beings. However, he categorically and repeatedly refuses to kill other Changers. Likewise, he will not impersonate other Changers either for reasons not explicitly revealed. The first time we learn about these principles is when he reminds the Idirans that "I've told you I won't impersonate another Changer. I certainly won't kill one" (Banks *Phlebas* 21). It is later revealed that Horza has killed two fellow Changers before on their home asteroid, Heibohre, and that this act, although it prevented a plot against the asteroid, saw him punished with temporary exile to an isolated world with meagre companionship. From the context of Horza's previous statement and his retelling of this event, it is clear that the killing (and perhaps impersonation) of other Changers by Changers is deeply taboo. Among his companions in his exile, we learn, is a female Changer, Kierachell, with whom Horza initiates a relationship. Of particular interest, however, is the fact that she is what is called a

"dormant Charger, one who had no training in and no desire to practise Changing" (Banks *Phlebas* 102). As one might expect knowing Horza fairly well at this point, Kierachell's and Horza's beliefs are almost diametrically opposed. Yet, "[she] had loved him, *body and mind*, despite it all" (Banks *Phlebas* 103; emphasis added). So, although the Changing seems as natural to Changers as any other bodily function, there is a minority of people who reject their Changer nature and instead chooses to live just like regular non-changing humans. Likewise, it is perhaps telling that the love, Kierachell holds for Horza is explicitly body and mind, when one considers that the housing of the mind, i.e., the brain, is the only immutable organ in a Changer's body.

The Culture Minds

Now, having located some particularities concerning the Changer culture and approach to their bodies, the thesis will now turn its attention towards the Culture, specifically some of the cultural complexities that make them so distinct in Banks' universe. Although, *Consider Phlebas* chiefly follows the viewpoint of Horza, a few chapters change the point of view to follow a different individual who is part of the Culture: the human genius Fal 'Ngeestra, and the recently constructed unnamed Culture Mind, chased by the Idirans, respectively. Along with these different points of view, some of Horza's observations and comments about the Culture will likewise be intercalated at times. As, although they are enemies, Horza devotes a considerable amount of his time on thinking about the Culture. For instance, his rationale for fighting the Culture, 'biological life', reveals some of his excessively prejudiced feelings about the Culture society. Particularly, his feelings about sentient machines are bordering on obsessive at times. For reasons, which will soon become apparent, we will now start at the beginning of *Consider Phlebas* to approach the sentient machines of the Culture.

The novel starts with a prologue that appropriately enough begins with the 'birth' of a Mind. The Mind itself is housed in a body described by the factory ship making it as: "it was a mongrel made from bits and pieces of different types of warcraft […] the dockyard threw the ship together as best it could from its depleted stock of components" (Banks *Phlebas* 3). It seems that there is a notion of body/mind even among the sentient machines of the Culture. This impression is further supported by Banks himself: "Culture starships - that is all classes of ship above inter-planetary - are sentient; their Minds […] bear the same relation to the fabric of the ship as a human brain does to the human body; the Mind is the important bit, and the rest is a life-support and transport system" (Banks "Notes"). Only one part of the ship is perfect: the new Mind residing inside the ship. Due to the nature of its construction the ship

does not have a name, and the factory craft will not give it one for two apparent reasons – lack of time, and the *real* reason: "the dockyard *mother* didn't give its warship *child* a name; it thought there was something else it lacked: hope" (Banks *Phlebas* 3; emphasis added). Shortly, afterward first the factory ship/mother, and subsequently the starship/body of the newly made Mind both perish, while the Mind itself manages to escape to Schar's World. This prologue is quite telling in how the Culture Minds think; while the Mind starship is constructed, it is also at the same time referred to as a child. The body of the Mind is not perfect, while the Mind itself is. The body in this case is unessential. Meanwhile, the factory ship refers to itself as the mother of the new Mind – it is not a creator or factory, but a mother.

Interestingly, names also seem to be something that have a certain gravity among the Culture Minds, which is something that Horza also ponders: "The Culture's General Contact Units […] had always chosen jokey, facetious names. Even now the new warships they were starting to produce […] favoured either jocular, sombre or downright unpleasant names" (Banks *Phlebas* 19). Examples of these names appear throughout the story: *Nervous Energy*, *Eschatologist, The Ends of Invention*, *Irregular Apocalypse*, *Profit Margin*, *Prosthetic Conscience*, *No More Mr. Nice Guy* and *Determinist*. With this selection of names, it is somewhat jarring that the unnamed Mind, which Horza is pursuing, ends up taking the name *Bora Horza Gobuchul* after the events of the novel. However, there does seem to be a logic behind the Mind choosing to take the name of its enemy. As Horza receives the news that his relationship with Yalson aboard the CAT has gotten her pregnant, despite them both believing such a coupling would be infertile, we also learn that Horza has no family; no one to carry on his name, and that this is something that greatly bothers him. Thus, Yalson proposing to keep the child, and let it carry on Horza's name, is a significant gesture for him. Unfortunately, Yalson and the unborn child die not soon after, leaving Horza once again with no one to carry his name. As Horza is lying close to death at the end of the novel, the significance of his name, his identity, is once again highlighted, as he cries out in terror asking what his name is (Banks *Phlebas* 441). At Balveda's reminder, he calms down and seems to welcome death. Accordingly, the Mind taking Horza's full conveys the impression of gratitude for saving it from the Idirans, and perhaps it seeks to honour his sacrifice by substituting Horza's unborn child by carrying his name, letting a memory of him live on.

Although, the previous paragraphs discussed the Minds, little knowledge of how they work, think or otherwise engage with humans has been presented so far, which is what the next few paragraphs will seek to rectify. Therefore, the perspective presented next will be

primarily that of Fal ‘Ngeestra, a human who works with the Minds of the Culture due in large part to her abnormal intellect, which astounds, confounds and fascinates even the super intelligent sentient machines (Banks *Phlebas* 87). Fal and her companion, the sentient Mind drone Jase, provide a semblance of nuance to the otherwise highly biased opinions of Horza.

Upon the first encounter with Fal and Jase, Fal’s musings about the drone, reveals two interesting titbits of information about the drone and to some extent its body. First, the drone is at least a millennium old, likely more, as: “for a thousand years or more Culture drones had had aura fields which coloured according to their mood – their equivalent of facial expression and body language” (Banks *Phlebas* 85). Jase being constructed before this technology was invented and then refusing to be refitted with the necessary components in the next millennium seems telling. Instead, it prefers to express itself by its tone of voice or, alternatively, to remain inscrutable. (Banks *Phlebas* 85). Jase’s refusal of these upgrades does indicate a certain aversion to changing its physical form and its way of communicating in general. Shortly afterwards, Jase reveals parts of its personality when it mulls over itself and considers itself a hopeless romantic, because it always records Fal’s laughs for itself, which it hides from her for a quite recognisable reason; it concedes to itself that, although, it knows sentient machines cannot die from shame, Jase believes that it would “do just that if Fal ever guessed any of this” (Banks *Phlebas* 89). When it secretly orders drones to bring Fal some pillows to support her since her leg is broken, Jase lies when Fal enquires if it did it, while being secretly pleased that she thinks so (Banks *Phlebas* 90). Albeit a minor character, Jase is an intriguing induvial, as it is the first and closest look at a sentient machine in *Consider Phlebas* at this point. Later, another sentient machine gets introduced, which we will return to, but first I want to discuss the Culture’s sentiment towards sentient machines further.

Life and Sentient Machines

These next paragraphs will be somewhat contradictory in that *Consider Phlebas* have two paradoxical views on the Culture, represented through Horza/Idirans’ negative sentiments and Culture/Fal/Balveda’s positive sentiments respectively, which this section will attempt to balance. Both perspectives tend to agree on the underlying facts of the Culture: machines are an integral part; humans are generally ‘freer’ and other notions such notions, but they tend to disagree on a basic, philosophical level as will become apparent.

Horza’s intolerance towards intelligent machines seems to be his reason for joining the war on the side of the Idirans against the Culture in the first place:

> At least they have a God, Frolk. The Culture doesn’t. […] They at least think the way you do. The Culture doesn’t. […] You want to know who the real representative of

> the Culture is on this planet? It's not her. […] it's that powered flesh-slicer she has following her everywhere, her knife missile. She might make the decisions, it might do what she tells it, but it's the real emissary. *That's what the Culture is about: machines*. You think because Balveda's got two legs and soft skin you should be on her side, but it's the Idirans, who are on the side of life in this war–. (Banks *Phlebas* 13; emphasis added).

As Horza clearly states, he is on the side of 'life' in this war as opposed to the side of the machines. It seems that the question of what constitutes life is of imminent importance for Horza and his reasons for joining the war effort. Conveniently, Horza later goes on to elaborate in detail on how exactly he defines this otherwise nebulous idea of what 'life' specifically is, and, furthermore, how being on the 'right side' justifies waging war: "I don't care how self-righteous the Culture feels, or how many people the Idirans kill. They're on the side of life – boring, old-fashioned, biological life; smelly, fallible and short-sighted, God knows, but *real* life. You're ruled by your machines. You're an evolutionary dead end." (Banks *Phlebas* 29). Clearly, Horza is exceptionally predisposed towards a certain interpretation of life. Horza's bias with sentient machines is not limited to his interactions with just Culture machines. When his starship, the *Clear Air Turbulence* - the CAT, accidentally picks up (kidnaps) the sentient non-Culture drone Unaha-Closp, Horza's treatment of it likewise betrays his xenophobic tendencies towards machines specifically: "[Unaha-Closp] would do as it was told, like machines always did. Only the Culture let them get so fancy they really did seem to have wills of their own." (Banks *Phlebas* 315). His treatment of Unaha-Closp is also something that seems to be starkly in contrast with what the drone itself considers fair, which we shall return to later. Likewise, the quoted line above reveal Horza's belief that only Culture Minds can be considered beings and not just machines.

Opposing this view on life and sentient machines' place in the universe are the Culture, perhaps most clearly demonstrated by this exchange between Fal 'Ngeestra and Jase discussing why the newly made Mind did not self-destruct: "it's called the instinct to survive […] it's programmed into most living things" (Banks *Phlebas* 93). Although, they do agree that exceptions to this instinct exists, the crucial detail here, is their argument that beings with a self-preservation instinct are living beings to begin with, whether organic or inorganic in nature.

Another interesting facet of this antagonism between the two stances towards the Culture is that of the role of sentient machines. Horza is convinced that the war between the Idirans and the Culture is the work of the Minds alone:

> He could not believe that the ordinary people in the Culture wanted the war, no matter how they had voted. […] The war had to be the Minds' idea; it was part of their clinical drive to clean up the galaxy, make it run on nice, efficient lines, without waste, injustice or suffering. The fools in the Culture couldn't see that one day the Minds would start thinking how wasteful and inefficient the humans in the Culture themselves were. (Banks *Phlebas* 35)

While initially this idea seems almost comically conspiratorial, there is some merit to Horza's musings, as we later learn from Fal: "[…] the Culture had placed its bets […] on the machine rather than the human brain. This was because the Culture saw itself as being a self-consciously rational society; and machines, even sentient ones, were more capable of achieving this desired state as well as more efficient […]" (Banks *Phlebas* 87). Not only has the Culture "placed its bets on machines", but it is also possible, Fal notes, to argue that the Culture is its machines – "that they represented [the Culture] at a more fundamental level than did any single human or group of humans within the society" (Banks *Phlebas* 86).

However, this is the most basic level on which the two agree; for while Horza assumes that the Minds will get rid of humanity at some point in the future, the reality, according to the Culture representatives, is different. The potential of the Minds leaves all the humans of the Culture with the chance to receive decent education, food and accommodations, along with the chance to pursue whatever activities they prefer (Banks *Phlebas* 87). The liberty of the Culture ties into two central precepts among its people, shared by both humans and Minds alike:

- "The only desire the Culture could not satisfy from within itself was one common to both the descendants its original human stock and the machines they had brought into being: the urge not to feel useless" (Banks *Phlebas* 451).
- "For the Culture's AIs, that need to feel useful is largely replaced by the desire to experience, but as a drive it is no less strong. […] the galaxy is, in other words, an immensely, intrinsically, and inexhaustibly interesting place; an intellectual playground for machines that know everything except fear and what lies hidden within the next uncharted stellar system" (Banks "Notes").

The fact that a few dozen people emerge in the human Culture population, who can match wits with the immensely intelligent Minds, is also a source of fascination to these machines (Banks *Phlebas* 87), and it must likewise tickle their fancy that through some truly astronomical numbers, they can experience somewhat intellectual equality with organic lifeforms.

Until now almost any mention of sentient machines in this part of the thesis, has been centred around Culture machines, but the universe of *Consider Phlebas* contains other sentient machines. One such machine even becomes a (reluctant) member of the CAT: the sentient drone Unaha-Closp, which will now briefly be discussed. Naturally, given Horza's prejudice against any kind of sentient machine the interactions between the two of them give rise to numerous clashes. Of these, one of the most meaningful is perhaps the one argument, in which Unaha-Closp feels provoked to explain its claim to personhood: "I would prefer, though, if you called me by my name, and not just by that word you manage to make sound like an expletive: 'machine'. I am called Unaha-Closp. […] I am not just a computer, I am a drone. I am conscious and I have an individual identity. Therefore I have a name" (Banks *Phlebas* 264). While it claims to be an individual, it is interesting that Unaha-Closp has had to have its sentience certified by an authority as an earlier argument with Horza illustrates: "I'll have you know I am an Accredited Free Construct, certified sentient under the Free Will Acts by the Greater Vavatch United Moral Standards Administration and with full citizenship of the Vavatch Heterocracy" (Banks *Phlebas* 260). No organic lifeform seems to have to go through the same troubles to be considered a sentient being in *Consider Phlebas*. This tirade is provoked by Horza's insistence to refer to Unaha-Closp as 'machine', a word and description it takes extreme displeasure in. In fact, it later equates being treated like a machine to be being treated like a slave (Banks *Phlebas* 407). Being able to discriminate between a machine and a sentient machine is not only important to a drone like Unaha-Closp, it is similarly necessary in the Culture, as Banks notes: "No machine is exploited, either; the idea here being that any job can be automated in such a way as to ensure that it can be done by a machine well below the level of potential consciousness" ("Notes"). Finally, the fact that Unaha-Closp features among the list of dramatis personae in *Consider Phlebas* also contribute to this impression of sentient machines being living beings and not just machines, as Horza so readily maintains.

***Software,* Analysis**

Having then investigated *Consider Phlebas,* the thesis will now change its focus to Rudy Rucker's novel, *Software,* which makes up the first novel in the Ware Tetralogy,

composed of *Software, Wetware, Freeware* and *Realware* respectively. Just like it was the case with Iain M. Banks' *Consider Phlebas*, a brief overview of the plot along with any other interesting data will accompany this introductory paragraph to this section of the analysis. With *Software* Rudy Rucker has the honour of being the first recipient of the Philip K. Dick award with the sequel *Wetware* also receiving a second Philip K. Dick award (Rucker "Wares"). Rudy Rucker himself is a mathematician and prolific writer of both fiction and non-fiction. I'll add one small note here talking about the Earth's natural satellite, the Moon. I choose to follow NASA's example of capitalizing the Moon when I refer to our moon as opposed to some style guides that prescribe referring to it in all lowercase even when talking about the Earth's natural satellite (David).

In *Software* the two protagonists are Cobb Anderson and Sta-Hi Mooney. Prior to the onset of the story, Cobb figured out how to give robots, so-called "boppers", free will and sentience. Now, at the end of his life, Anderson is invited to the Moon, where the boppers live, to receive immortality by the grateful machines. Sta-Hi tags along to the Moon. The boppers on the Moon are currently engaged in a civil war, which Sta-Hi inadvertently joins on the side of the small boppers. Meanwhile, Anderson's consciousness is 'brain-taped', and he is subsequently downloaded into a robotic body made in his likeness. The story ends with Sta-Hi and Anderson's standoff on Earth and in aftermath Anderson's consciousness storage hardware is destroyed and his consciousness is likewise erased for good.

Structurally much like *Consider Phlebas*, there are several areas of interest to us in *Software*, which will be individually tackled. Specifically, three aspects are going to be the chief concern:

- Cobb Anderson, the creator of the boppers.
- Sta-Hi and his interactions with and opinions about the boppers.
- The characteristics of the boppers themselves

The next few paragraphs will attempt to treat each of three points above separately, but inexorably some overlap might occur, although attempts will be made to limit any such overlaps.

Cobb Anderson

Beginning then with Cobb Anderson, he is arguably the creator of the intelligent/sentient robots of *Software*; he even goes so far as to consider himself the father of the first group of intelligent robots (Rucker 72). When the reader first meets Cobb Anderson, he is an old man, living on borrowed time with his cheap replacement heart. The boppers

want to offer him immortality, because of Cobb Anderson helping them achieve sentience and letting them live/"let[ting] the robots get out of control" as his fellow humans insist (Rucker 21). The mechanic copy of Cobb, dubbed $Cobb_2$, even suggests that they can supply him with enough tank-grown organs to rebuilt him completely (Rucker 17). This lie is one the boppers keep up almost right up until they are going to put Cobb through his immortality 'operation', before which Cobb realises that the operation is not a replacement of his organic parts, but rather quite the reverse where Cobb's *software*, his brain and mind/consciousness, will be extracted and uploaded. Thereby, the boppers can then transfer his consciousness into any synthetic body they would want. Conveniently, there is already a copy of his body on Earth in the form of $Cobb_2$. This process is, as Cobb ironically remarks akin to suicide: "*I'm committing suicide to keep from getting killed*" (Rucker 75; emphasis in original).

From this point on Cobb is no longer an organic being, on the contrary, he at this point inhabiting mechanic bodies with his mind being kept on tape. After his successful operation, he finds himself in the $Cobb_2$ body back on Earth. Here, he first struggles with the fact that he is suddenly living in a body with memories of his neighbour moving in and initiating a relationship with her, while it did not happen to *him*. Shortly, afterward they are to attend a prom together, and it is here some of the thought-provoking facts of his new existence become clear to us. First, he struggles with the habit of meeting his basic human needs: consuming foods and liquids. Exceptionally, pertinent is the fact that Cobb, who has been living in an alcohol induced stupor for years suddenly cannot get drunk. He finds a letter addressed to him by the boppers wherein they provide him useful information for his new life:

> 1) Your body's skeleton, muscles, processors, etc. are synthetic and self-repairing. Be sure, however, to recharge the power-cells twice a year. Plug is located in left heel.
> 2) Your brain-functions are partially contained in a remote supercooled processor. Avoid electromagnetic shielding or noise-sources, as this may degrade the body-brain link. Travel should be undertaken only after consultation.
> 3) Every effort has been made to transfer your software without distortion. In addition we have built in a library of useful subroutines. (Rucker 97)

There is no doubt that Cobb is now a synthetic being. Before continuing, it is prudent to account for exactly what the distinctions between 'software' and 'hardware' are in *Software*. Conveniently, Cobb provides this exposition next:

> Intellectually he had always known it was possible. A robot, or a person, has two parts: hardware and software. The hardware is the actual physical material involved,

> and the software is the pattern in which the material is arranged. Your *brain* is hardware, but the *information* in the brain is software. The mind... memories, habits, opinions, skills... is all software. (Rucker 97-98)

Accordingly, it is only one half of Cobb that is still the original: the software, although no longer stored in the organic brain, it should still be the same Cobb. This is, initially at least, not the immortality Cobb had in mind (Rucker 98). The new body's subroutines, however, reveal functionalities that far surpass a normal, organic human body's functionality though, and it would perhaps make some people apprehensive to have direct control over some of them: "'MISTER FROSTEE, TIME-LINE, ATLAS, CALCULATOR, SENSE ACUITY, SELF-DESTRUCT, REFERENCE LIBRARY, FACTCHUNKING, SEX, HYPER ACTIVITY, DRUNKENNESS... '" (Rucker 98; emphasis in original). Nonetheless, the revelation of these subroutines gives Cobb something to take pleasure in with his new body as we shall now see.

With the effect of the DRUNKENNESS subroutine, Cobb goes out with Annie, and he starts to consider that "the whole situation didn't seem so horrible and frightening as it initially had. Hell, he had it made" (Rucker 106). Although, as he considers, the DRUNKENNES-effect, is not quite the same as the real thing, it is convenient that he can change his level of inebriation with just a breath or two. There is one problem for Cobb at the prom though: spending time with friends serves as a powerful reminder that he is now a robot: "But, listening to his friends talk, he had a feeling of shame at no longer being human", and Cobb starts to consider whether he is now "guru or golem?" (Rucker 108). However, the shame of being synthetic is short lived for later Cobb claims: "Cobb hadn't enjoyed himself so much in years" (Rucker 113). Unfortunately for Cobb, the peaceful evening is interrupted by Sta-Hi and Mooney, which prompts him to consider that "what the boppers had done to him was, on the whole, a good thing" (Rucker 115), and he unsuccessfully tries to convince Mooney and Sta-Hi that everyone should get a "nice everlasting body like mine" (Rucker 116). And here the crux of two incompatible perspectives materialises, which we will recognise as belonging in the transhuman camp or rejecting it, respectively:

> "It's not so unreasonable," Cobb protested. "It's a natural next evolutionary step. Imagine people that carry terabyte computing systems in their head, people that communicate directly brain-to-brain, people who live for centuries and change bodies like suits of clothes!"
>
> "Imagine people that aren't people," Sta-Hi replied. "Cobb, the big boppers like TEX and MEX have been trying to run the same con on the Moon. And most of the little

> boppers up there aren't buying it... most would rather fight then let themselves be patched into the big systems. Now why do you think that is?"
>
> "Obviously some people... or boppers... are going to be paranoid about losing their precious individuality," Cobb answered. (Rucker 116)

This struggle between the big boppers and little boppers and humans wanting to keep their individuality is the central conflict of this novel, and it is a topic that will come to the forefront of the bopper section. For now, the thesis will turn its attention to Sta-Hi.

Sta-Hi

Sta-Hi is interesting as a counterpart to Cobb Anderson because they are at quite contrasting in their interactions with the boppers. Cobb invents and helps the first boppers with their uprising happening in 2001, when Sta-Hi was only six years old. In a way, Cobb and Sta-Hi's respective backgrounds make them both particularly biased regarding the boppers. Cobb sees himself as their benevolent father while Sta-Hi has been fed the generally accepted human consensus that free boppers are deeply undesirable, starting the moment his parents throw his Ralph Numbers toy out the window of their car, when they learn of the rebellion (Rucker 72). It follows then that the propaganda that Sta-Hi has been exposed to leaves him with certain prejudices against the boppers, which will now be examined.

Sta-Hi's first experience with the intelligent robotic boppers is meeting his own robotic copy, dubbed Sta-Hi_2 (Rucker 45). It seems then that his double takes pleasure in Sta-Hi's shock. Sta-Hi's next encounter with a bopper is somewhat more ambiguous for the man himself. Shortly after arriving on the Moon, Sta-Hi learns that the stewardess from their flight is in fact both a bopper and a human at the same time. However, she is perhaps more robot than human as Sta-Hi finds her manner strange when he run into her at the hotel: "Alone in a booth at the end of the room was the face he wanted. The stewardess. There was no drink in front of her, no book... she was just sitting there. […] blank as a parked car" (Rucker 60). The juxtaposition of Misty's human appearance and her robotic behaviour is a source of continuous confusion to Sta-Hi, and it seems that the explanation of how exactly one can define Misty is not opaque either. Misty used to be a human, but her brain was forcefully taped and thus one of the big boppers now has a perfect copy of the original Misty's personality which can then be used to remote control the mechanic copy of her body that the boppers made. But while Misty is part of DEX, the big boppers' memory banks, she also appears to have free will:

> "I'm surprised you're telling me all this," Sta-Hi said finally. […] "But BEX didn't want me to," Misty was saying. "*You* can't hear him of course, but he's been telling

> me to shut up the whole time. But he can't make me. I still have my free will... it's part of the brain-tape. I can do what I like." She smiled into Sta-Hi's eyes. There was a moment's silence and then she started talking again. "You wanted to know who I am. I gave you one answer. A robot-remote. A servo-unit operated by a program stored in a bopper spaceship. But... I'm still Misty-girl, too. The soul is the software, you know. The software is what counts, the habits and the memories. The brain and the body are just meat, seeds for the organ-tanks." (Rucker 62)

This conversation reveals several interesting aspects of *Software*. Firstly, the brain-taped humans are both a part of the boppers while also being independent from them. Free will appears to be a natural consequence of them having their own software. Secondly, the argument, which Misty makes here, is that her intrinsic humanness is not a result of her physical makeup, but rather it is a consequence of her being born as, living as and experiencing life as a human, which makes her the human being, Misty.

While the distinction between 'human' and 'machine/robot' makes perfect sense for Misty, who lives this reality, the explanation, as it is provided to Sta-Hi, fails to elucidate on the contrast, as he notes when they engage in sex:

> "The sex was nice, but confusing. The whole situation kept going di-polar on Sta-Hi. One instant Misty would seem like a lovely warm girl who'd survived a terrible injury, like a lost puppy to be stroked, a lonely woman to be husbanded. But then he'd start thinking of the wires behind her eyes, and he'd be screwing a machine, an inanimate object, a public toilet." (Rucker 62)

Moreover, it seems that while the distinction between treatment of machines and humans is clear for Sta-Hi: *inanimate object* and *warm girl*, Cobb is better prepared to appreciate that the boundary between them is less unambiguous as his actions in their shared hotel room display. When he realises that he is prancing around naked in front of Misty, Cobb tries to cover up as best as possible. Meanwhile, Sta-Hi questions his actions, referring to Misty as "just a robot-remote" (Rucker 63), which Cobb simply ignores. Him treating Misty as a person and not an inanimate object incites an argument between Misty and Sta-Hi, in which Misty questions exactly what Sta-Hi deems the boundary between human and machine: "'Why did you say *just* a robot-remote? As if I were less than human. Would you say that about a woman with an artificial leg? Or a glass eye? I just happen to be *all* artificial.'" (Rucker 64; emphasis in original). Cobb, a short while later, queries Sta-Hi why he would send Misty away. It is, at this point in the novel, obvious that Sta-Hi is a product of a society that abhors and considers boppers and likewise the humans, who have been brain-taped, little

more than machines. This notion is additionally illustrated in the last confrontation between Sta-Hi and Cobb.

Although, Sta-Hi has helped the regular boppers blow up GAX, a big bopper, and he has been part of symbiosis with a bopper, Sta-Hi refers to as ‘Happy Cloak’, his views are still incompatible with Cobb Anderson when they meet face to face after Cobb’s taping:

> "Now that you think you're immortal you don't worry about death," Sta-Hi said bitterly. "That's really enlightened of you. But whether you know it or not, Cobb Anderson is *dead*. I saw him die, and if you think you're him, you're just fooling yourself." He sat down, suddenly very tired. "If I'm not Cobb Anderson, then who would I be?" The flickercladding face smiled at him gently. "I *know* I'm Cobb. I have the same memories, the same habits, the same feelings that I always did."
> "But what about your... your *soul*," Sta-Hi said, not liking to use the word. "Each person has a soul, a consciousness, whatever you call it. There's some special thing that makes a person be alive, and there's no way that can go into a computer program. (Rucker 133)

Before engaging in this (failed) dialogue, there is a moment where Sta-Hi’s supposition above shines through: “The robot began to talk then, slowly, and in Cobb's old voice” (Rucker 133). While Sta-Hi abstains from directly debating the finer points of when a human becomes too artificial for him with Misty, it seems here that Sta-Hi has found out precisely what his objection towards human/bopper interface is: that of the nebulous concept of soul. In fact, so nebulous a concept is it that Sta-Hi himself cannot seem to pin it down himself, only that there is an intangible discrepancy between a bopper and a human. Sta-Hi is prepared to partition a machine with an (almost) perfect visually replicated body; with a copied memory; habits and feelings of a human being from the original organic human being. It seems a singularly limiting way of defining a concept of *humanness* from everything else, which is an idea the discussion will return to. Now, the third identified subject of interest in *Software* will be addressed – namely the boppers themselves.

The Boppers

The original purpose of the boppers was to mine and work on the Moon. Cobb Anderson manages after a while to give the boppers intelligence, which soon results in the 2001 revolt, after which the boppers are in control of the Moon, and its production facilities. One of the subjects that these next few paragraphs will concern itself with is describing the boppers themselves for, if they can be considered people, it will be interesting to notice how they diverge from other beings (humans). Save for the Sta-Hi_2 and $Cobb_2$ copies, the reader’s

first encounter with the boppers occurs on the Moon where Ralph Numbers and Wagstaff are conversing with each other in binary code, or as the boppers refer to it: “sacred binary bits of machine language” (Rucker 27). This reveals fascinating insight into the culture of the boppers. First, the boppers consider binary a language for communicating in – in this case like English. Second, however, is the addition of the adjective: *sacred*. Ralph Numbers clearly differentiates the functions of binary with English, when he switches from binary to English, once he realises that Wagstaff, and he are going to argue. In fact, the boppers s seem to have a spirituality to them: they refer to the shelter of ‘the One’ as a “modernistic church. Which, in some sense, it was” (Rucker 29) - a sentiment shared or even embellished by the big bopper, Mr. Frostee: “Simpler beings merge to produce higher beings, and they must merge and merge again. In this way we draw ever closer to the One.” (Rucker 119). With this statement, and the discussion that follows between Cobb and Mr. Frostee, the purpose for taping both human beings and little boppers alike by the big boppers is revealed. While Cobb is not immediately in agreement with Mr. Frostee, his exchange with Sta-Hi later displays his changed perspective: “All consciousness is One. The One is God. God is pure existence unmodified.” (Rucker 133). The boppers’ spirituality is of interest while the humans of *Software* are plainly not pious to any extent. They are, however, alien to humans in other ways, which will be the focus of the next paragraph.

As has already become apparent earlier, the boppers are undoubtably able to create bodies that look like human bodies, but in *Software* it is readily apparent that they choose not to make these bodies except to accommodate the brain-taped humans. Instead, the actual boppers’ appearances are as diverse as one can imagine. Ralph Numbers and Wagstaff, who were mentioned previously, are the first boppers that are described explicitly in *Software*, and they are not alike in appearance at all: Ralph is described as looking like a file cabinet on legs with five arms and a few visible lights. The sentences describing him also remark that: “it was hard to tell what he was thinking” (Rucker 27). Wagstaff, meanwhile, is described “much more expressive” with an appearance that is slightly reminiscent of St. Georges dragon (Rucker 27). Boppers’ bodies themselves are also distinctly unlike human bodies in that they are specifically built. A lot of the bopper industry on the Moon is devoted to the production of bopper components since they are all hardcoded to allow the One to erase them every ten months without fail. Every single bopper is therefore on the clock to build a copy of themselves, a *scion*, for when they are inevitably wiped by the One. Although the consciousness of the boppers is taped into the new scion, there is still the inescapable truth that a 100% perfect copy is impossible. After one of bopper factions, the diggers, assassinate

Ralph, he is taped into his scion, but the fact remains that Ralph still has to get used to his new scion: "No two arrangements of circuit cards can be *exactly* the same, and adjusting to a new body takes a while […] It was like putting on a new pair of glasses, only more so." (Rucker 47). By building this scion themselves, the boppers also have a chance to change their physical shape. Ralph Numbers, specifically, does, according to his own musings, not like to change his physical appearance too much even though it appears that a makeover would be a relatively short endeavour (Rucker 48). Unique appearances appear to be an intrinsic part of bopper culture, for as Sta-Hi remarks up landing on the Moon: "Sta-Hi had seen models of a few of the basic types before, but no two of them waiting out there looked quite alike" (Rucker 56). Although, the bopper body is to some degree interesting, there are other aspects of boppers that also high light their personhood, namely their emotional lives.

Some of the complex emotional responses of the boppers have already been discussed previously. One such emotionally loaded scene is the one, in which Ralph Numbers questions Cobb Anderson if he knew that he would disobey and begin the revolt that would lead to the boppers' freedom (Rucker 72). Similarly, the quarrel between Wagstaff and Ralph deteriorate until it is clear that they are opposite sides of the coming conflict. Yet, Wagstaff finds himself unable to "attack so great a bopper at close range" (Rucker 29). Clearly, the two examples here showcase a capacity for mutual respect of other beings. Equally interesting is the revelation that boppers, mechanical beings, can feel "something like a sexual love for each other" as Ralph and the bopper Burchee have (Rucker 49). They have apparently "conjugated several times" (Rucker 49), according to Ralph; a process that for boppers presents itself as being perhaps even more intimate than it can ever be for humans since their processors merge completely during the 'act'. Lastly, the boppers are obviously able to have their own opinions about the world they find themselves in, as evidenced by the fact that the plotting of the big boppers is leading the little boppers to unite against them in what amount to civil war between the two factions.

With this close examination of Cobb Anderson, Sta-Hi and the boppers themselves, it is now time to discuss the findings from both *Consider Phlebas* and *Software*, and finally to discuss the similarities and differences in these two texts.

Sentient life and Humanity, Discussion

Having analysed and extracted parts of Rudy Rucker's *Software* and Iain M. Banks' *Consider Phlebas*, the following section will undertake a discussion of the extracted data and treat it in relation to the theoretical work, previously elucidated on. The objective at the end of this section is to compare the two texts both in terms of contrasts and similarities.

In these next few paragraphs, the focus will specifically be on the consciousness and self, and how the existence of self is no longer intricately tied up with consciousness. Further, Foerst's notions of emotion being a necessary part of cognition and her ideas on personhood will be tied into this part of the discussion. Beginning then with *Consider Phlebas*, the consciousness of a being is a diverse thing. The text deals with many types of beings – both biological and nonbiological in origin. Specifically, we see a people like the Changers, where their bodies are changed almost exactly according to their consciousness with the sole exception being their inability to change their own brains, the wetware on which the software of consciousness is running much a like a computer program. On the other hand, a being like the non-Mind drone Unaha-Closp must have its own sentience certified, even though a sentient machine is nothing extraordinary. The certification will also not change Unaha-Closp's feelings about itself, it is merely a formality that reinforce its pride as a being. Finally, beings like the Culture Minds are so alien to most humans that describing their consciousness is a fool's errand. Hayles' assertion that a posthuman subject is also a post-conscious subject (280), is perhaps not immediately true. Here it is obvious that to talk about consciousness in any way in relation to *Consider Phlebas* is much more easily done by approaching it at a broader and more intangible level than it already is when just dealing with Homo Sapiens, and perhaps disregarding consciousness completely as a requisite for self is most natural at that point.

The paradigm shift in defining cognition and consciousness that Foerst mentions broadens and simplifies the requisites for the aforementioned terms to include emotions. In Jase and Fal's conversation along with the introduction with the Culture factory ship, we see plenty of evidence to support that the sentient machines have emotions. Jase is a hopeless romantic; every being has an instinct to survive, and the factory ship lacks the hope to even name its 'child'. This also ties into Foerst's thoughts on defining personhood and her insistence on adopting a broad, inclusive interpretation of personhood. Again Unaha-Closp's certification of sentience is interesting. Likewise, the fact that the Culture specifically goes out of its way to avoid the exploitation of and fair treatment of any sentient being, organic or inorganic alike, seems to be a recipe for success – at least in-universe of *Consider Phlebas* as the Culture ends up defeating the Idirans, whom it might be argued, are on the side of God and 'natural' life.

Consciousness in *Software*, however, is fittingly enough far more susceptible to direct intervention as we see. Both humans and robots alike in the novel must live with the fact that consciousness can be taped and stored within the memories of the big boppers, and thus I can

also be erased as is the case with Cobb Anderson. The most immediate example of this is of course the 'immortality' deal Cobb Anderson agrees to. While his thoughts and mannerism are presumably unchanged after his successful taping, a process that invokes Hans Moravec's imagery to a frightening degree, the nature of his new existence also means that while he can still act with free will, the big bopper Mr. Frostee can forcefully eject Cobb from the body he is inhabiting. Likewise, the forced erasure of the boppers every 10 months also raises questions about consciousness and self. For while Ralph Numbers has undergone this erasure successfully 36 times, can a 'self' even exist if your entire consciousness is erased and then a copy of that taped into a new, almost copy of your body? In this case, it is likely most prudent for anyone to consider the existence of self as separate from consciousness. The personhood of the brain-taped humans is also in question through all of *Software*. Cobb's initial confusion with his new existence after his taping is one clue. The bigger indicator is, however, Sta-Hi's attitude towards Cobb when they meet towards the end of the novel, where he accuses Cobb of being a robot that thinks it is a human, and not the opposite, i.e. a human in robot body. For Sta-Hi the process that tapes the brain, inevitably results in that person losing their 'humanness'. The problem is almost a textbook definition of Foerst's discussion on personhood and having a narrow definition of what makes a person. Likewise, the boppers themselves are also exempt from being considered beings on equal footing to humans with only Cobb considering them as his children and treating them with kindness and understanding, which they reciprocate by treating him as closer to an equal than the poor humans, they essentially kidnap and murder to tape their brains.

Next, I wish to discuss the *body/embodiment* duality in *Software* and *Consider Phlebas*, starting again with the latter text. Obviously, the question of *body*/*embodiment* in relation to the Changers is a difficult question, but also one that I expect can be a clue for a potential answer to why the Changers as a people are so universally abhorred by the different humanoid species. The Changer *body* by itself appears to be uniquely Changer while still looking human. As an example, the outworld minister body that Horza has assumed when we first encounter him is old. So, while Horza himself is not actually old, he has a body that is in fact that of an old man. A fact that continues to be a hindrance to him until he can fully Change again. Therefore, it is hard to talk about a normative body when considering Changers. However, the embodiment of Changers is a bit of a conundrum for embodiment as it has been described by Hayles. For according to her, embodiment is intricately tied to context and the specific individual; it follows then that it is also profoundly personally connect to each individual. The problem naturally arises then, when you have a species that

can mimic people to such an extent that they can fool their fellow crewmates or co-worker ministers. As we see, the Gerontocrat Amahain-Frolk almost refuses to believe that Horza is not in actuality the man, whose appearance he has assumed. This leads me to state that the extreme human aversion to Changers can be explained by the fact that they readily ignore the implicit personalised embodiment with impunity. This transgression of implicit almost universal human existence is met with huge mistrust as we learn. Ironically, for Horza (and perhaps the rest of the Changers) the society most accepting of their unique bodies and lackadaisical approach to the sanctity of embodiment would be the human-machine society of the Culture, where the transgressive nature of Changing would not be as obviously unique. Since the Culture humans have the technology to change gender, it is reasonable to assume that they perhaps have a more accepting attitude towards the question of embodiment and transgressing it. Likewise, the society of the Culture is also made up of a large percentage of sentient machines, which presumably would not care as much about the Changers abilities. Lastly, it is telling that some Changers themselves choose to deliberately neglect their natural abilities, perhaps feeling some form of aversion to taking on another's embodiment, or maybe feeling unsure about their own embodiment at the thought of being somebody else?

While *Software* does not have shape changing humans on the level of the Changers, it too challenges views on embodiment. However, the notion of 'the body' is also challenged or even disregarded to an extent in *Software* when considering the boppers. Through the descriptions of Ralph Numbers, Wagstaff, the big boppers, who are a hotel, a spaceship, a museum and a factory respectively, and the other small boppers a picture starts to emerge of a culture that does not care about possessing a normative body at all. In fact, the notion of a normative body among the boppers in antithetical. Their way of referring to body parts, or specifically in their case as components, also reveals a disregard for a normalised body. The robot-human question is, however, less clear cut to understand. The argument between Sta-Hi and Misty for example raises a valid question in terms of the 'body' and being human. For where does the normalisation of bodies begin to become problematic in terms of artificial limps – one limp, two, eyes, most of the internal organs? When does the body cease to be human and become machine? Does it even matter? It clearly matters to humanity in *Software* considering the lengths the boppers and the human-robots go to keep their existence a secret. Sta-Hi himself is also deeply confused and even disturbed by Misty. One moment he is talking to or having sex with a human girl, the next her gestures or actions remind him that it is just a machine he is engaging with. Since this body/embodiment duality no longer serves its purpose for Sta-Hi, he instead chooses to disregard what he sees or experiences, and he

doubles down on the fact that humans have some intangible essence that robots and machines lack – in his case he tentatively identifies it as the soul, which is echoed by other characters in the novel. However, for Cobb and others the soul is the software, and it is therefore not lost upon transfer between the brain and the bopper. It is perhaps the heart of the conflict between human and bopper acceptance.

Cobb Anderson is uniquely situated in the story as both a mouthpiece for humanity and later the robot-bopper-humanity that is emerging thanks to the boppers' taping of human brains. Through him, we see the process of him getting used to his new body – which can perhaps more than anything be considered a normalised version of 'Cobb Anderson', as it will remain unchanging unless he wishes a change. We see him initially struggle with the fact that he now has a body that needs no sustenance and in particular alcohol. It seems, in fact, that Cobb's drinking has become a deeply habitual pattern in his life, and until the discovery of his new *subroutines*, his enjoyment of robot life is limited. Afterwards though, we quickly see him adjusting his embodied habits to the new life. He swiftly becomes proficient at keeping his 'fake' intoxication going, and later his exploration of the SEX subroutine also seems to have become almost instinctual for him. As the story proceeds, we also see Cobb becoming more open to the idea of him modifying his body at will; the thought of potentially changing into a woman intrigues Cobb. As a cult leader this idea of changing the body at will becomes reality, and his behaviour has likewise changed to match closer to that one would expect from a cult leader to match.

Finally, I will now directly discuss and compare *Consider Phlebas* and *Software*. On the surface, the two texts do not appear to resemble each other much. *Consider Phlebas* is a space opera set in a universe much more advanced than ours, populated by numerous types of humans and sentient machines with some of those being hyperintelligent. Moreover, it also features a race of pious aliens. *Software,* on the other hand, with its cyberpunk motifs is immediately more recognisable for the reader. It takes place on Earth and the Moon, and while the technology in *Software* is more advanced, it still somehow appears more within the realms of possibility. Still, the two texts share similarities in that the existence of artificial machine-based life are established in both stories, and in meeting with humanity it brings strife. I have predominantly discussed three basic notions pertaining to posthumanism in this section of the thesis: consciousness, personhood and body/embodiment. Consciousness is not as distinctly articulated in *Consider Phlebas* as it is in *Software*, where it is a driving force in the plot, but there are still grounds to discuss consciousness as being less important. In fact, it seems to be in contrast with *Software* where consciousness is often likened to a 'soul' that

even sentient machines do not possess. This depersonalisation of sentient machines is a recurring theme in both works, where it seems that the characters in both texts predominantly discard the notion of sentient machines as being people on the grounds of their 'unnatural' origins, i.e. synthetic existence.

The body/embodiment topic is also of immense interest in both texts as people in either novel have rather outlandish options when it comes to changing their bodies. This of course also means that the notion of a normalised 'body' becomes a battlefield, where the question of what constitutes 'humanness' is all-encompassing. Any human form that goes beyond normal limits of body/embodiment are seen as transgressive, and as such are subjected to extreme prejudice by their fellow human beings. Predictably, it is the artificial lifeforms in these novels that have the most relaxed and perhaps even indifferent attitude towards questioning body/embodiment notions. It is telling that the constructed lifeforms are laissez-faire towards these things with their largely unchanging bodies, while the organic lifeforms are much more aware of and affected by these ideas. It certainly paints a picture of synthetic lifeforms being more predisposed towards posthuman ideals while the humans struggle more to reconcile their reality with the posthuman notions listed here. Kelly Hurley in her reading of posthuman identity in *Alien* and *Rapid*, makes the observation about quasi-human figures' representation: "Such posthuman embodiments are liminal entities, occupying both terms (or rather, existing in the slash between them) of the opposition human/not-human" (203), suggesting that there is no definitive answer to what a posthuman body is, since it is in a constant state of change. Perhaps this is the key to accepting bodies that are not quite normalised human bodies?

Conclusion

The thesis set out to try to examine the following statement:

- *How is the interrelation between humanity and sentient machines presented in Banks' Consider Phlebas and Rucker's Software; how can posthumanism specifically be used to analyse this; and finally, how do these two texts approach the question of humanity and synthetic life co-existing?*

In the course analysing and discussing the two texts certain parallels as well as contrasts show up; the notion of *the body* and *embodiment* as well as *consciousness* in *Consider Phlebas* and *Software* are under scrutiny and perhaps even in the middle of an evolution. Both texts provide amble evidence to support that a humanist centric view is dissatisfactory to account for the changes happening in the books, and more importantly a humanist centric viewpoint is anathema to peaceful coexistence between human and synthetic life. The

boundary between human and machine is especially circumspect in *Software*. Still, many of the conflicts between the synthetic lifeforms and human beings inhabiting either of the texts arise from limiting what personhood is: the sentient machines wish to be treated like living beings rather than things, and often the humans of the stories will repudiate this desire. At the same time, it seems that the society of the Culture is profiting particularly from a broad definition of living being by being equally accepting towards different humans and sentient machines.

The theoretical framework and likewise the discussion would have been bolstered by supplementing with another posthumanist theorist's framework to encourage a greater range of discussion and a greater understanding of posthumanism as the one employed by N. Katherine Hayles has foundations within the cybernetics field of study. Someone like Donna Haraway, whose theoretical work springs from a feminist background, would perhaps have provided valuable insight which could have strengthened and supplemented the arguments in the thesis. Particularly, the interactions between the characters Misty and Sta-Hi in *Software* could have formed the basis for an interesting feminist analysis.

Similarly, the notion of consciousness is too nebulous of a concept in the way that Hayles uses it. However, even attempting to account for consciousness/intelligence/cognition in any satisfactory manner would have resulted in a thesis that would have varied distinctly from this one. Still, it is interesting to consider whether humanity feels threatened by the irrefutable emergence of a consciousness that is not our own. Already today, questions are being raised about the ethical nature of octopus farming due to their intelligent nature (Marshall). So, the question of perceived human superiority due to consciousness is particularly relevant now. Still, I will echo Iain M. Banks' assertion that the emergence of a sentient artificial lifeform would be a paradigm shift for humanity, for we would no longer be alone to face the rest of eternity, and from the point of emergence and onwards our future would be intrinsically tied to that of synthetic lifeform whether through coexistence or disharmony – and would that not be interesting times to live in?

Works Cited

Asimov, Isaac. *I, Robot*. Harper*Voyager*. 1950.

Banks, Iain M. "A Few Notes on the Culture". *Vavatch Orbital*. Originally posted to *newsgroup rec.arts.sf.written*, 1994, www.vavatch.co.uk/books/banks/cultnote.htm. Accessed 8 Dec. 2021.

--- *Consider Phlebas*. Orbit. 1987.

Bolter, J.D. (2016). Posthumanism. In The International Encyclopedia of Communication Theory and Philosophy (eds K.B. Jensen, E.W. Rothenbuhler, J.D. Pooley and R.T. Craig). DOI: https://doi-org.zorac.aub.aau.dk/10.1002/9781118766804.wbiect220

"Consider Phlebas" *Iain-Banks.net*. n.d., www.iain-banks.net/titles/iain-m-banks-3/consider-phlebas/9780748109999/. Accessed 18 Dec. 2020.

@daisyowl. "if you ever code something that "feels like a hack but it works," just remember that a CPU is literally a rock that we tricked into thinking." *Twitter*, 15 March 2017, 5:03 PM, https://twitter.com/daisyowl/status/841802094361235456

David, Leonard. "A capital debate: Should Earth's natural satellite be 'Moon' or 'moon'?" *Space.com*, 4 Mar. 2020, www.space.com/moon-name-controversy-capital-letter.html. Accessed 22 Dec. 2020.

Flasiński, Mariusz. *Introduction to Artificial Intelligence*. 1st ed. 2016., Springer International Publishing, 2016, doi:10.1007/978-3-319-40022-8.

Foerst, Anne. "Artificial Sociability: from embodied AI toward new understandings of personhood". *Technology in Society*, vol 21, issue 4, 1999.

Foster, Thomas. "Cybernetics and Posthumanism". *A Companion to Literary Theory*, D.H. Richter (Ed.), 2018. https://doi-org.zorac.aub.aau.dk/10.1002/9781118958933.ch36

Gibson, William. *Neuromancer*. Ace Science Fiction Books, 1984.

Halberstam, Judith & Livingston, Ira. Introduction. *Posthuman Bodies*, edited by Judith Halberstam & Ira Livingston, Indiana University Press, 1995, pp. 1-19.

Hassan, Ihab. "Prometheus as Performer: Toward a Posthumanist Culture?". Vol 31, No. 4 Winter. The Georgia Review, 1977. URL: https://www.jstor.org/stable/41397536.

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

Hurley, Kelly. "Reading Like an Alien: Posthuman Identity in Ridley Scott's Alien and David Cronenberg's Rabid" *Posthuman Bodies*, edited by Judith Halberstam & Ira Livingston, Indiana University Press, 1995, pp. 203-224.

Jockers, Matthew L. *Macroanalysis: Digital methods and literary history*. University of Illinois Press, 2013.

Lettvin, Jerome Y., et al. "What the frog's eye tells the frog's brain." *Proceedings of the IRE* 47.11. 1959: 1940-1951.

Marshall, Claire. “The world’s first octopus farm – should it got ahead?” *BBC news*, 20 dec. 2021, www.bbc.com/news/science-environment-59667645. Accessed 1 Jan. 2022.

Maturana, Humberto R. & Varela, Francisco J. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company, 1980.

Novikov, D. A. *Cybernetics: From Past to Future*. Vol. 47. Cham: Springer International Publishing AG, 2015.

Ray, Thomas S. “An Evolutionary Approach to Synthetic Biology: Zen and the Art of Creating Life”. *Artificial Life* 1, no. 1/2, 1993.

Rubin, Charles T. “Artificial Intelligence and Human Nature”. *The New Atlantis* no. 1 (spring), 2003

Rucker, Rudy. *Software*. Prime Books. 1982

Squier, Susan M. “Reproducing the Posthuman Body.” *Posthuman Bodies*, edited by Judith Halberstam & Ira Livingston, Indiana University Press, 1995, pp. 113-132.

Turing, Alan M., “Computing machinery and intelligence”, *Mind*, vol. LIX, Issue 236, October 1950, Pages 433–460. https://academic.oup.com/mind/article/LIX/236/433/986238. Accessed 22 December 2020.

Umbrello, Steven. “Posthumanism.” Con Texte (Sudbury), vol. 2, no. 1, 2018, pp. 28–32, https://doi.org/10.28984/ct.v2i1.279.