

**Artificial Identity: Representations of Robots and Cyborgs in
Contemporary Anglo-American Science Fiction Films**

By

Em C. Gibson

A thesis submitted in partial fulfilment for the requirements of the degree of MA (by
Research) at the University of Central Lancashire

January 2012

ABSTRACT

The 'human condition' has traditionally been an area of study addressed primarily by philosophers concerned with the mind/body problem, rather than studied as a neuroscientific conundrum. However, contemporary developments in science and technology that afford us a greater knowledge of the human brain have resulted in an increased scientific focus on consciousness, emotion and personhood. This thesis argues that such explorations into consciousness and emotion as prerequisites of 'artificial identity' have entered the domain of contemporary cinema through the representations of robots and cyborgs.

Despite the capacity for transhumanist practice and the creation of artificially intelligent automata that these developments have made possible, blurring the line between organic human and mechanical robots, it remains common for no distinctions to be made between the terms 'human' and 'person', which are used interchangeably to describe a member of the human race. Philosopher Daniel C. Dennett, though, has proposed a series of criteria for personhood that challenge the assumption that only humans can be considered persons.

The application of his criteria to a series of key texts that highlight the relationships between humans and representations of automata - *I, Robot* (2004, Dir. Alex Proyas), *Terminator Salvation* (2009, Dir. McG) and *Bicentennial Man* (1999, Dir. Chris Columbus) – is central to this thesis. It explores the extent to which the representations of robots and cyborgs can be considered persons within utopian and dystopian narratives that have, at their core, a view of artificial identity as desirable or as nightmare.

In conjunction with Dennett, the theories of neurologist and neuroscientist Antonio Damasio are applied, which explore both the biological means by which emotional (rather than solely physical) feelings are generated in humans, and the capacity of humans to simulate emotion. As Damasio argues that many of the central operations of the human central nervous and visceral systems are reducible to fundamental physics, the suggestion is that robots, too, could also 'experience' consciousness and emotion, being as they are very simplistic versions of humans. As such, the application of these theories suggests that the representations of robots and cyborgs in the key texts could be considered persons.

TABLE OF CONTENTS

	Page
INTRODUCTION:	
Aims.....	1
Selecting the Key Texts.....	2
Film and Science.....	3
Representations of the Future: Fears and Dreams.....	6
Transhumanism.....	10
The Non-Human Person and Criteria for Personhood.....	13
CHAPTER ONE:	
<i>I, Robot</i> : Techno-Utopianism.....	20
Detective Spooner.....	21
Dr. Susan Calvin.....	24
Sonny.....	29
CHAPTER TWO:	
What is it to be Human?	35
Views and Approaches.....	36
Determinism Versus Dualism.....	39
Rejecting Free Will.....	41
Emotion and Reason.....	46
The Somatic Marker Hypothesis: Does Consciousness Always Result in an Emotional Response?	49

CONCLUSION.....	56
Bibliography.....	61
Filmography.....	65

ACKNOWLEDGEMENTS

Firstly, I wish to acknowledge the help and support that I have received from my supervisory team. As Director of Studies, Dr. Helen Jones has given me invaluable support, advice and patience not only throughout the writing of this thesis, but since I began my first degree, for which I am truly grateful. My sincere thanks to my second supervisor and 'bad cop', Paul Humble, for sharing his wealth of knowledge and steering me in the right direction when I needed it, and my appreciation also to Professor Ewa Mazierska and Dr. Johan Siebers for their time, advice and guidance.

Secondly, I owe my love and thanks to my family, to whom I am indebted for their constant presence and support, without which I would not have attempted this thesis. Especially Scott, who has always been there pushing me and picking me up without being asked; my Mum, who has read more essays than either of us care to remember, and my Nan, Stefan, Cameron and Mars, who together provided the inspiration for this thesis. Thank you all. Lastly, my gratitude and admiration to John and Melendy, who never let me down.

INTRODUCTION

AIMS

It is the aim of this thesis to investigate contemporary filmic representations of artificial identity by drawing in cognitive and philosophical studies using the work, among others, of philosopher Daniel C. Dennett and neurologist and neuroscientist Antonio Damasio. It is posited that traditional binary opposites in the representations of humans and robots in audio-visual science fiction are being eroded in contemporary popular Anglo-American films. Consequently this thesis seeks to identify to what extent these automata may now be considered persons, and what the difference is between a person and a human person.

Human knowledge of recent scientific discoveries and technological developments that make the line between cyborg and human often more difficult to distinguish, have resulted in a stronger theoretical focus on the emotional rather than solely cognitive aspects of brain function, explored increasingly in contemporary science fiction film narratives that emphasize the relationships of the human protagonists with representations of automata. Such developments support this study, which links cognitive theory in film studies research with philosophy.

Accordingly the work of Dennett is relevant, which explores the possibility of robotic consciousness, arguing that the belief that robots might one day develop consciousness is reasonable because ‘...we human beings are conscious, and we are a *sort* of robot ourselves’ (2009, p.186). Though Dennett’s work is drawn from philosophy rather than film studies and based on existing automata rather than fictional versions, it is still clearly applicable to the representations of robots and cyborgs in recent filmmaking as the images I will be studying are, in part, based on or inspired by the robots Dennett discusses.

Further, Dennett offers a specific set of criteria for personhood which will be applied to the representations of robots and cyborgs in the selected texts studied herein, to determine whether or not they may be termed persons, despite not being human. As the texts have both human and robotic protagonists, the qualities displayed by the human characters in addition to those in Dennett’s criteria for personhood may serve to highlight the differences between a person and a human person.

Antonio Damasio’s work is relevant to the study of these narratives: it takes a neurological perspective and is not a critique of science fiction films, as it addresses the ability of the human brain to experience emotions and the conditions necessary for the creation of those emotions. Nonetheless, it is applicable to a discussion that compares human and robot identity and the

capacity of the latter to develop consciousness – one of Dennett’s criteria for personhood. For example, he argues that: ‘...rationality does not function without emotional input. Rationality stems from our emotions, and emotions stem from our bodily senses’ (2006, p. 67). The representations of automata in the science fiction narratives selected are represented as developing bodily senses, albeit generated by comparatively crude inorganic sensors. Hence the films seek to suggest that there is the potential for them to experience some level of emotional experience.

The scope that is achievable in a thesis of this length has inevitably led to omissions. For example, the textual analysis focuses on mise-en-scene, dialogue and plot development; it does not address the use of music or sonic effects that help to identify the film’s mood as utopic or dystopic and provide further clues about genre and characterisation. Although numerous contemporary science fiction film adaptations, including those studied here, address the issue of ‘otherness’, this thesis also omits discussion of race, focussing instead on the differences between persons and human persons. So, for example, film references to skin and its degeneration as a signifier of the mortal – and by implication non-robotic – body are relevant; black skin as an external signifier of race – and hence of difference within the human – is beyond the scope of the thesis. Finally, whilst I make reference to Holland’s text *Descartes Goes to Hollywood* I do not pursue her much stronger focus on gender. Instead, I address the extent to which the representations of robots and cyborgs in the key texts can be considered persons, or human persons, outside of the gender debate.

SELECTING THE KEY TEXTS

The key texts that form the body of this work are *I, Robot* (2004, Dir. Alex Proyas), *Bicentennial Man* (1999, Dir. Chris Columbus), and, to a lesser extent, *Terminator Salvation* (2009, Dir. McG). Each film was selected in part for their contemporary nature to highlight the current level of erosion of a long-standing binary opposition: for example, *Bicentennial Man* charts the journey of an NDR-114 robot from home help to human. The robot, who is named Andrew, and who was originally purchased by Richard Martin (Sam Neill) to complete the household chores, spends his entire existence with Martin’s family, watching them grow up around him, and eventually die. As he watches their development, their eventual passing, and the reactions of those left behind to their loss, Andrew ponders what life would be like as a human, and begins his quest to become one in a narrative that spans two centuries.

In *Terminator Salvation*, the artificial intelligence network Skynet is co-ordinating a massive assault of machines on humankind in a post-apocalyptic 2018. Only John Connor (Christian Bale) and his small band of resistance members stand between the machines and total annihilation of humans. But when a stranger, Marcus Wright (Sam Worthington), appears, the clear-cut binary opposition of man versus machine is thrown into question through gradual revelations of his past as a human on death row, who left his body to scientists experimenting with cyborg technology.

Of the three films, the greatest attention is paid to *I, Robot*, given its central exploration of differences between 'them' and 'us', robot and human. *I, Robot* is an exploration not just of the binary opposition between human and robot upon which this thesis is built, but of humanness itself, and the identification therefore of Dennett's 'special ingredient'. In the film, set in Chicago, 2035, Detective Del Spooner (Will Smith) is an emotional, reactive police officer with an inherent distrust of all things robotic. When respected scientist and creator of robots Dr. Alfred Lanning, with whom we later discover Spooner has a complex past, suddenly commits suicide, the officer sets out to investigate what he quickly comes to believe is something far more sinister than an old man's choice.

Chapter One of the thesis presents a close study of Spooner, Dr. Susan Calvin (Bridget Moynahan) and the robot Sonny (Alan Tudyk), and argues that whilst there is no attempt by Proyas to suggest that the robots in this text are in any way human, the application of Dennett's criteria for personhood reveals that Sonny exhibits a range of behaviours sufficient for him to be categorised as a person. Therefore, comparisons between Sonny and the human characters, Spooner in particular, are drawn in order to ascertain the differentiating factor/s between persons and human persons.

FILM AND SCIENCE

Films over time have represented areas of science to differing degrees of accuracy. As scholarly subjects, though, Film Studies and Science have remained separate because the consumer-driven nature of the texts studied in the former tends to relegate exact representations of science to inconsequential levels. The visual spectacle and entertainment required to make such texts financially viable has often taken precedence instead.

In his article on the links between science and film, Christopher S. Rose explains that movies '...representing science are often at odds with each other simply because entertainment and

education are two different things.¹ To me, this brings to mind Antonio Damasio's argument that, contrary to long-held popular beliefs, emotion and reason not only can co-exist, but are interdependent on one another.² In this sense then, the entertainment of films represents the emotion, and the factual basis of science, the reason.

Although filmic texts still need to be entertaining according to the conventions of their given genre, and any critical analysis of the themes of such texts relies on the individual decoders that form a heterogeneous mass audience comprising many cultures and beliefs, nonetheless these films do represent society and the many scientific and technological changes within it. Often, for example in the case of the cyborgs and robots upon which this thesis is built, filmic representations are the closest many members of the public will get to these scientific advances. As a result, an increasing number of writers like Rose have begun to take a cross-disciplinary approach to the study of film, science and philosophy, placing my work as part of a growing body by film critics, scientists and philosophers who 'draw connections and parallels between entertaining movie science and exciting real world science.'³

As our technological and scientific abilities increase, the possibilities for the future that such changes make possible – both utopian and dystopian alike – are inevitably explored in films. These recent 'rapid advances in biotechnology' and '...their implications for personhood' are discussed by Bill Muehlenberg.⁴ Though he approaches the topic from a religious standpoint and thus ultimately contrasts said developments with traditional Christian teachings of creation by a single God, his explanation of the resulting potential threats to our humanity and loss of personhood nonetheless raise the need to define not just the person, but the human person, in the face of a techno-scientific landscape that now enables us to create human clones, human-animal chimeras, cyborgs and artificially intelligent robots.

The biotechnical themes that Muehlenberg calls 'an assault on personhood'⁵ he notes are present in, amongst others, *I, Robot*, *Bicentennial Man* and *Terminator Salvation*, confirming these as appropriate texts through which to study the representations of robots and cyborgs in terms of their personhood and resulting artificial identity. His work also serves to exemplify the

¹ Rose, C. S., 2007, *Biology in the Movies: Using the Double-Edged Sword of Popular Culture to Enhance Public Understanding of Science*, *Journal of Evolutionary Biology*, Issue 34, p.49, in <http://resources.metapress.com/pdf-preview.axd?code=3016t2154360t6m5&size=largest>, accessed 31st December 2010.

² Damasio, A., 2006, *Descartes' Error*, Preface, p. xxii.

³ Ibid.

⁴ Muehlenberg, B., 2008, *Bioethics, Hollywood and the Biblical Worldview*, in *Culturewatch*, <http://www.billmuehlenberg.com/2008/07/17/bioethics-hollywood-and-the-biblical-worldview/>, accessed 30th December 2010.

⁵ Ibid.

link between neurobiology and the humanities – a ‘utopian bridge’ envisioned by Antonio Damasio in 2005, that could ‘explore the threads that interconnect neurobiology to culture’.⁶

Similarly, Samantha Holland discusses the trend seen in contemporary films to address Cartesian philosophy, specifically the mind-body issue, and personal identity, noting the use of ‘images of the technologised body to investigate questions of self-hood ...and the threats posed to such concepts by postmodern technology and A.I.’⁷ In the same text, Holland highlights the permeation of what she describes as our ‘cyborg fascination’ into other media – comics, video games, toys and novels. Thus her work both highlights, and is an example of academic study which links film, philosophy and science.

In addition to Damasio, Holland, Muehlenberg and Rose, Professor Patricia Pisters has published on film-philosophical questions such as those addressed by Holland, and co-organised an international conference in Amsterdam, 2010, where she gave a presentation entitled *The Brain is the Screen: The Neuro-Image in Contemporary Cinema*,⁸ which drew on films including *Fight Club* (1999, Dir. David Fincher) and *Eternal Sunshine of the Spotless Mind* (2004, Dir. Michel Gondry) to highlight the methodological connections between art, science and philosophy. From a philosophical stance, Pisters argues that ‘movies can modify our subjectivities such that the brain and mind are one.’⁹

In their chapter entitled *What is it to be Human? Blade Runner and Dark City*,¹⁰ Deborah Knight and George McKnight again highlight the link between the three disciplines through their discussion of the mystery of identity presented in science fiction classic *Blade Runner* (1982, Dir. Ridley Scott). Scott’s film, like *Bicentennial Man*, studies the title question from the point of view of a replicant living in a dystopic future, as opposed to the comfortable existence of Andrew. Accordingly, the earlier film presents a replicant living a human life as a crime, whilst in the later text, humanity becomes Andrew’s ultimate goal, perhaps symptomatic of the technological advancements and increased awareness of the public of such technology in the intervening period.

⁶ Damasio, A., 2006, *Descartes’ Error*, Preface, p.xx.

⁷ Holland, S., 1995, *Descartes Goes to Hollywood: Mind, Body and Gender in Contemporary Cyborg Cinema*, in *Body & Society Online Journal*, Vol. 1, No. 3-4, p.157.

⁸ Pisters, P., 2010, *The Brain is the Screen: The Neuro-Image in Contemporary Cinema*, in <http://www.sbs.ox.ac.uk/centres/insis/neurosociety/Documents/pisters-neurosociety.pdf>, accessed 31st December 2010.

⁹ Pisters, P., 2006, *The Spiritual Dimension of the Brain as Screen Zigzagging from Cosmos to Earth (and Back)*, in <http://home.medewerker.uva.nl/p.p.r.w.pisters/bestanden/Screen%20Consciousness%20article%20as%20published%20met%20referentie.pdf>, accessed 31st December 2010.

¹⁰ Knight, D., and McKnight, G., *What is it to be Human? Blade Runner and Dark City*, in Sanders, S.M., (Ed.), 2008, *The Philosophy of Science Fiction Film*, Ch.1, p.21-37

Other relevant theorists include Roz Kaveney, who, in reference to *Strange Days* (1995, Dir. Kathryn Bigelow), discusses a filmic representation of 'the effects of technology on social and personal life'¹¹, and says (of *Total Recall*, 1990, Dir. Paul Verhoeven) that once viewers willingly enter a constructed world, such as those in the key texts upon which this thesis is based, and '...interact with beings that believe themselves to be real, the status of your own knowledge of your own world comes up for grabs.'¹² In the same book, Kaveney also uses the examples of C3PO and R2-D2 in the *Star Wars* films to consider the potential of a robot or cyborg to teach a person how to become more human and in doing so, become more human themselves.

Though Kaveney's work does not explicitly question what it is that determines human identity and thus what differentiates human protagonists from their robot or cyborg counterparts, her discussion of C3PO's flawed masculinity and role as a mentor raises the issue of identity in science fiction films not just in terms of the humans in these narratives, but also the artificial identity of the robots and cyborgs concerned. Similarly, in *The Philosophy of Science Fiction Film* (2008), editor Steven M. Sanders claims on the dust jacket that science fiction films 'provide a medium through which questions about personal identity, moral agency, artificial consciousness and other categories of experience can be addressed.'

Jason Holt picks up on these ideas with regards to both artificial intelligence and artificial consciousness in his work. His study of the *Terminator* franchise, for example, provides an important historical context to the proposed study of *Terminator: Salvation* which had not been released at the time of Holt's writing. Representations of artificial identity clearly alter across the film quadrilogy and the thesis will refer, for example, to the case of Marcus Wright in *Salvation* who in the narrative was, and still believes, that he is human.

REPRESENTATIONS OF THE FUTURE: FEARS AND DREAMS

Every film is a documentary. These films give tangible expression to our wishes and dreams, our nightmares and dreads. They make the stuff of imagination concrete – visible and audible. They give a sense of what we wish, or fear, reality itself might become.¹³

Film is a means of transmitting cultural codes and values, thus texts commonly reflect the culture and ideology of their producers. Science Fiction, a genre synonymous with metallic robots, fantastic future technology and extra-terrestrial beings, both friend and foe, is no different. Due

¹¹ Kaveney, R., 2005, *From Alien to the Matrix: Reading Science Fiction Film*, p.85.

¹² *Ibid.*, p.65.

¹³ Nichols, B, 2002, *Introduction to Documentary*, p1.

perhaps to its far-fetched nature, imagined and often impossible or implausible situations and characters, alongside the occasionally under-developed production values and tendency to quickly date of many filmic examples, Science Fiction is often viewed as the preserve of the geek. As a result, Science Fiction can suffer from reluctance on the part of society to place the same value and status on all but the most shining, epic, mainstream narratives as tends to be afforded to other genres.

However, fictional tales are undoubtedly popular with all age groups, genders and cultures. Regardless of the budget, and however unrealistic these far off planets and intergalactic narratives, from *Metropolis* (1927, Dir. Fritz Lang) to *Moon* (2009, Dir. Duncan Jones), Science Fiction continues to attract a worldwide audience seeking escapism or insight, who want to immerse themselves in the possibilities and predictions that the genre offers. In contrast to numerous other genres, Science Fiction's frequent focus on the future rather than past or present makes it an arena in which we may explore our beliefs, uncertainties and hopes for events not yet existing. In the same way it may be argued that Disaster Films in many ways represent the fears and interests of the societies producing them regarding what has happened or is happening, so Science Fiction represents our feelings towards events yet to happen.

Importantly though, the title *Science Fiction* is somewhat misleading in the sense that it does not address the elements of the genre that represent real, non-fictional events, or at least are based upon them. Bill Nichols categorises films into two distinct types, which he defines respectively as fiction and non-fiction: '...documentaries of wish-fulfilment' and '...documentaries of social representation.'¹⁴ Science Fiction is a combination of both fictional and non-fictional elements – narratives drawn from the imagination that 'offer worlds for us to explore and contemplate', along with '...truths, insights and perspectives we may adopt as our own or reject.'¹⁵

The fictional elements may represent either our dreams or dreads depending on the context of the film, but it is the representation of macro-social issues and micro-social interpretations that make Science Fiction appealing and 'give tangible representation to the aspects of the world we already inhabit and share.'¹⁶ Through the analysis of the aesthetics and interaction of the principal characters in a number of contemporary Anglo-American Science Fiction films, this thesis seeks to determine the extent to which the representations of non-human characters may be considered persons. The application of Dennett's criteria for personhood will be fundamental in determining this, whilst Damasio's Somatic Marker Hypothesis will play a significant role in

¹⁴ Ibid.

¹⁵ Ibid.

¹⁶ Ibid.

recognizing both the similarities and differences between human persons and non-human persons. Such differences enable the identification of Dennett's 'special ingredient', and may form the basis of the contemporary dreams and nightmares that Nichols refers to.

Audio-visual representations of advanced technology, including a range of automata and cyborgs is nothing new to an audience who first experienced these characters in their technological infancy (though arguably this would have appeared less so given the historical context in which they were made) with Fritz Lang's false Maria from *Metropolis* (1927) or Robby the Robot from *Forbidden Planet* (1956, Dir. Fred M. Wilcox). The likes of C3PO and R2-D2 from *Star Wars* (1977, Dir. George Lucas) still captivate and entertain audiences internationally, as does James Cameron's more recent mechanical assassin in *The Terminator* (1984). Similarly, it is commonplace for humans to wear devices like pacemakers, or to have prosthetic limbs.

However, the social and technological context within which the audiences who experience these texts live is constantly evolving. With the increasingly common use of 3D technology in cinemas and 4D technologies on theme park experiences, even the arena in which the text is viewed is becoming ever more advanced. It is that social and technological context that alters the dreams, nightmares and perceptions of reality and potential futures which are then represented in Science Fiction narratives.

In 'Half Man, Half Machine: The Cyborgs Are Coming'¹⁷, Pat Pilcher outlines the dreams that current developments in technology may make a reality. 'Until now, man-machine hybrids, or cyborgs, have been the stuff of trashy sci-fi flicks', says Pilcher, in an article based on implantable electronics. However, recent developments in silk-silicon technology, which allow nanometre-thick silicon circuits to be implanted into the human body without rejection, have opened up a wide range of possibilities.

Tattoos that may be used for anything from blood sugar monitoring to advertising become possible. As does the notion of an interface between the human nervous system and these electronic circuits, providing a direct link between humans and the technology around them, that may be of direct benefit to disabled members of society. Even more impressively, this new technology opens up the potential for inroads to be made into the treatment of degenerative conditions like Parkinsons disease and the revolution of prosthetic limb design. Nichols' ideas surrounding 'wish-fulfillment' are simple to address here.

¹⁷ Pilcher, P., 2009, *Half Man, Half Machine: The Cyborgs Are Coming*, in <http://www.independent.co.uk/life-style/gadgets-and-tech/features/half-man-half-machine-the-cyborgs-are-coming-1822739.html>, accessed 5th Jan 2010

However, with great power comes both great responsibility and the opportunity for that technology to be misused. Pilcher cites a range of possible misuses of human embedded technology, from the everyday office prank – making your boss kick themselves – to techno-terrorism, and the ‘would-be cyber assassins’ capable of shutting down a pacemaker. Similarly, Johann Hari again discusses filmic references to advanced automata, arguing that ‘the idea of machines that are designed to whirr out into the world and make their own decisions to kill is an old sci-fi fantasy’, however ‘...in a world of such whooshing technological transformation ...the concept has leaped in just five years from the cinema screen to the battlefield.’¹⁸ Hari cites the 12,000 robots on active duty in Iraq as evidence of this, and a US governmental report detailing both the current reliance on ‘killer robots’, developed by British laboratories, and the likelihood of autonomous robots being ‘the norm’ on battlefields in the near future.

Of course, such technology again raises issues when control is lost. When autonomy is then built into such robots as a means of combating the loss of control, for example to insurgents in Iraq, the irony of this action becomes apparent. Humans seek to develop themselves and evolve beyond natural means, becoming transhuman. We strive to develop ever more complex robots and technology, yet losing control of that development and the ability of others to misuse it has always been a source of the nightmares Nichols discusses. Hari goes on ‘...robots find it almost impossible to distinguish an apple from a tomato: how will they distinguish a combatant from a civilian? You can't appeal to a robot for mercy; you can't activate its empathy.’ With any scientific evidence to the contrary rarely if ever shown in Science Fiction texts, issues like this become and common fodder for us to explore in the cinematic arena.

Science Fiction narratives, particularly those that are animated, appeal to the younger viewer – the spectator of the future – and initiate them into the conventions of the genre and present those same dreams and nightmares to be considered from an early age. Of course, given the youth of each new generation of viewers, these issues are dealt with in a manner generally less complex than is the case with those texts targeted at an older audience. The full potential impacts of the issues raised are often not focussed upon, in favour of a more optimistic, if perhaps somewhat unrealistic restoration of equilibrium. Nor do the origins of the represented situations tend to be

¹⁸ Hari, J., 2010, *The Age of the Killer Robot Is No Longer a Sci-fi Fantasy*, in <http://www.independent.co.uk/opinion/commentators/johann-hari/johann-hari-the-age-of-the-killer-robot-is-no-longer-a-sci-fi-fantasy-1875220.html>, accessed 5th Jan 2010.

addressed in detail, despite the potential positive impact of such a representation, but nonetheless the issues are raised.

Visionaries (1987, Sunbow Productions) presents a futuristic utopia where man is served by machines, which is thrown into chaos when those machines become aware and begin to enslave their human creators.¹⁹ *Wall-E* (2008, Dir. Andrew Stanton) charts the journey of a small robot tasked with cleaning up the waste and by-products of our consumer-based society, and our current pre-occupation with furthering our own evolution, transforming ourselves into something bigger, better and stronger is reflected in the plethora of superhero-with-alter-ego tales. *He-Man and the Masters of the Universe* (1982, Filmation) in particular is a useful study here due to a range of incarnations spanning three decades.

The original 1980s version shows Adam use magic to transform into the most powerful man in the universe, aided by highly skilled warriors and advanced science. Planetary exploration and the possibility of being overrun by technology are also explored. By the time of the arrival of the latest version, *Masters of the Universe vs. The Snake Men* (2002, Mike Young Productions), targeted at a slightly older audience, magic remained but there was a renewed focus on technology and its interaction with the human body, representing technological developments in society in the intervening period. For example, one episode²⁰ shows an alien being provide He-Man with a suit of 'battle armour' extremely reminiscent of that worn by Ellen Ripley in *Aliens* (1986, Dir. James Cameron).

TRANSHUMANISM

Philosophical questions about what defines both personhood and human personhood have faced new challenges recently as a result of advancements in biotechnology, such as cloning and innovative prosthetics, which have lead to the identification of a transhumanist position. That being a position supported by individuals who advocate morphological freedom, and the incorporation or use of advanced technology to augment and improve the human body, so as to reduce or eliminate our existing physical and mental limitations, such as illness, aging, physical weakness and death. Defined by Nick Bostrom:

It holds that current human nature is improvable through the use of applied science and other rational methods, which may make it possible to increase human health-span, extend

¹⁹ *Visionaries Knights of the Magical Light*, 1987, Episode 4: *The Price of Freedom*, Sunbow Productions.

²⁰ *Masters of the Universe vs. The Snake Men*, 2002, Episode 33: *Of Machines and Men*, Halperin, M., original air date 29/11/03, Mike Young Productions.

our intellectual and physical capacities, and give us increased control over our own mental states and moods. Technologies of concern include not only current ones, like genetic engineering and information technology, but also anticipated future developments such as fully immersive virtual reality, machine-phase nanotechnology, and artificial intelligence.²¹

With acknowledgement given to a range of potential hazards and objections to such beliefs, Bostrom continues that ‘transhumanists promote the view that human enhancement technologies should be made widely available, and that individuals should have broad discretion over which of these technologies to apply to themselves.’ Ultimately, he claims that ‘human enhancement technologies will offer enormous potential for deeply valuable and humanly beneficial uses that [...] may make us, or our descendants, “posthuman”.’²²

However, the implications of transhumanist practice, which may lead to a perceived or actual shift in human nature and corporeality towards beings like the robots and cyborgs studied in this thesis, leads to bioconservative objections to the ‘dehumanizing’ qualities of the transhumanist position. Typically, these fall into one of two categories – ethical issues or practical issues. In his book *Enough: Staying Human in an Engineered Age* (2003), Bill McKibben argues against transhumanism from an ethical standpoint. It should be noted here that, as with many of the future technologies represented in the key texts studied in this thesis, such potentially groundbreaking scientific advances inevitably entail an incompleteness of knowledge of the practicalities and implications of such developments, which in itself is an ethical issue when the application of these technologies is considered. Further, physical constraints such as access to advanced technologies present additional ethical issues which films like *Gattaca* (1997, Dir. Andrew Niccol) are often used to highlight.

One significant ethical criticism of transhumanist practice is the perceived threat it brings to human values. For McKibben, it is the trivialisation of human identity and the removal of natural forces and constraints on human existence, discussed earlier, which either provide or enhance the meanings of everyday practices. Alongside concern for the emotional side of human existence, lies the Asimov-termed ‘Frankenstein complex’,²³ effectively depicted by *The Young Family* (2002-3) by Patricia Piccinini.²⁴

Shelley’s novel is frequently referenced by critics of transhumanism, particularly in light of recent scientific and technological breakthroughs in cloning and genetic engineering. In line with this,

²¹ Bostrom N., 2005, *In Defense of Human Dignity*, p.1, in <http://www.nickbostrom.com/ethics/dignity.html>, accessed July 2011.

²² Ibid.

²³ Asimov, I. 1978, *The Machine and the Robot*, cited in Warrick, P.S., Greenberg, M.H. and Olander, J.D. (Eds.), 1978, *Science Fiction: Contemporary Mythology*, p.137.

²⁴ Piccinini, P., 2002-3, image at <http://www.patriciapiccinini.net/wearefamily/index.php?sec=yf>, accessed 31st December 2010.

Keekok Lee raises concern over the transformation of the 'natural' to the 'artefactual'²⁵, citing the replicants in *Blade Runner*. The implications here on free will (discussed later), assuming you accept its existence, are significant. For example, how would a modified human determine if their successes or failures, let us say, in a sporting arena, were down to their own training and diet choices, or simply due to their augmentation? Depending on the answer, questions may also be raised as to how free any resulting actions are.

In response, transhumanists assert that chimeras, clones and other products of the transhuman process may not be human, but still persons worthy of respect, and that it is hypocritical to make judgements of transhumanism based on ethics, when the identification of such creations a 'the other' in a negative sense by humans is unethical in itself. It is not clear what term transhumanists would afford to humans who have augmented themselves technologically to go far beyond what those we currently term human can do. Are they super-human, and we merely human? Are they something completely separate, or are they the new human, and we by comparison now sub-human?

The intention of this thesis is not to study and promote either side of the transhumanism debate. However, the issues that transhumanism raises are directly relevant to a study of representations of robots and cyborgs, the concept of personhood, and, though it is not the focus of this thesis, the societies which produced the filmic texts in which these appear. I have already introduced Nichols' ideas of films as 'documentaries of wish fulfilment' and 'social representation', therefore the transhumanist movement and technological developments in society such as those discussed in relation to Hari and Pilcher, are likely to have had an impact on these films, and those produced in the future.

As increasingly artificially intelligent systems are created, alongside synthesised (as opposed to naturally born) biological life, what criteria do we use to determine at what point personhood is achieved by such creations, if at all? If we are to define personhood, how do we distinguish then between being a person, and being a human person, as several theorists (discussed later) argue that there are many non-human persons? And, given our ability to synthesise and replace human organs with both organic (human and non-human) and inorganic alternatives, after how many

²⁵ Lee, K., 1999, *The Natural and the Artefactual: The Implications of Deep Science and Deep Technology for Environmental Philosophy*, Ch. 2, p.50, in http://books.google.co.uk/books?id=ZeH62s45c3oC&pg=PA49&lpg=PA49&dq=The+Natural+and+the+Artefactual&source=bl&ots=KQ-PMNYvFc&sig=d0qZ6cSSG7T7me8F3LMJ0vJe4nY&hl=en&ei=1Z-8S6WqEY3QmgPT0tnCCA&sa=X&oi=book_result&ct=result&resnum=1&ved=0CAgQ6AEwAA#v=onepage&q&f=false, accessed 7th April 2010.

organ replacements might an individual, considered a human person, cease to be a human person, or indeed a person at all?

Is it the number of replaced organs that determines the answer to such a question, whether or not the replacements are organic, or is it instead which specific organs are replaced? As the brain is often hailed as the key to our individuality as humans, if a human brain was transplanted into another body, be that organic or inorganic, would that new 'host' be considered human, and would the 'donor' lose their humanity, despite being otherwise human by birth? The regular citation of filmic texts such as those that form the focus of this thesis in answer to debates such as this, alongside the work of numerous theorists including those discussed here, place this thesis as part of an ongoing debate about the nature of personhood, and the representation of personhood in films.

Though this thesis is not specifically concerned with legal definitions and rights pertaining to persons, as we are defining personhood it is worth noting that the law makes room for corporate personhood as well as that of the individual. The latter will be the focus of this thesis, with the acknowledgement that an argument could be made for the corporate personhood, or even personhood as an 'individual' of US Robotics in *I, Robot*, and its counterparts in the other key texts studied herein.

THE NON-HUMAN PERSON AND CRITERIA FOR PERSONHOOD

The terms 'human' and 'person' have been used 'interchangeably' throughout history, notes Thomas I. White in his text *Dolphin People* (2010)²⁶, which examines the possibility of personhood in non-human animals, using dolphins as the core species studied. As a result, he argues, many people are surprised by or even reject the possibility that any being not considered a human could still be termed a person. As those beings referred to tend to be organic, the likelihood of the robots and cyborgs studied in this thesis being considered persons is reduced based on White's thoughts, particularly in the case of Sonny and Andrew who are robotic both in appearance and construction.

He continues to describe how members of the human race have believed themselves to be part of a unique species with 'advanced intellectual and emotional traits', so much so '...that we use 'human' and 'person' as synonyms', and view animals as less important than humans, despite our membership of the animal kingdom. However, White claims that 'the idea of non-terrestrial

²⁶ White, T.I., 2010, *Dolphin People*, in *The Philosopher's Magazine*, Issue 49, 2nd Quarter 2010, p.36-43.

intelligence is no longer in the domain of science fiction.’ A philosophical distinction may be made, he says, between the two terms, placing humanness within the bounds of science, and viewing ‘person’ as a philosophical concept.

This allows the representation of robots and cyborgs in this thesis to be addressed from both points of view – scientific and philosophical – and provides a basis for a distinction to be made between the automata and the definition of ‘human’ that science offers. For White, ‘...‘human’ refers to any member of the biological category homo sapiens.’ That is, they are born into a specific species. ‘Person’, he continues, ‘refers to the combination of advanced traits by which we define ourselves.’ Later, I will present arguments from other theorists who typify these ‘advanced traits’ in a more detailed manner and who note consciousness as key to personhood. By extension then, they suggest that it is possible for someone biologically human to not be designated as a person; for example those in a persistent vegetative state.

Given White’s distinction between the two conditions, Hayward, Cardinal and Jones’s argument that the ‘one characteristic necessary’ for personhood is ‘membership of the human race’,²⁷ I find immediately problematic. They acknowledge complexities such as emotions – fear and excitement specifically – in animals as cerebrally simplistic as chickens, but quote Jean Paul Sartre’s view that ‘Man is, indeed, a project which possesses a subjective life’²⁸ in support of their argument which refuses to grant intelligent animals such as dolphins and chimpanzees any degree of personhood.

Although they offer a seemingly compelling point in support of this, regarding the greatly reduced (or in the given example of swotting a fly, non-existent) penalties for harming an animal compared to those for abusing a human, this point rests on legal rights, rather than on the status of a non-human animal’s personhood. Their assertion is that a single characteristic, membership of the human race, is necessary to be considered a person. Yet despite the importance they place on this condition, they fail to define what this membership requires.

If to be ‘human’ and thus to have membership is to be entirely naturally born, comprising of all of the correct bodily parts a human should have, what about amputees and those born with physical defects including malfunctioning or missing organs, or absent or additional limbs, as in the case

²⁷ Hayward, J., Cardinal, D., and Jones, G., *Are There Any Non-Human Persons?* in http://www.alevelphilosophy.com/index.php?option=com_content&view=article&catid=54%3Aessays&id=48%3Aare-there-any-non-human-persons&Itemid=66, accessed 8th Jan 2011.

²⁸ Ibid.

of Lakshmi Tatma, born with a parasitic twin attached?²⁹ The impact of a persistent vegetative state or mental defects to the extent that membership of the human race is impaired or impossible is not considered. Numerous cases of feral children like Genie who, when discovered was incontinent, unable to chew, cry, barely able to speak, walk, swallow or properly focus her eyes³⁰, was in behaviour and cognition more akin to a non-human animal or perhaps even a robot, demonstrate the need to effectively define what characteristics are necessary for an individual to attain the membership that Hayward, Cardinal and Jones discuss.

Conversely, in *All Animals Are Equal* (1989), Peter Singer argues that although there are key differences between humans and non-human animals – cerebral capacity for example – which lead to the differences in rights that Hayward, Cardinal and Jones note, this is ‘no barrier to extending the basic principle of equality to non-human animals’. Typically, he claims, most philosophers attempting to identify the differences between humans and non-human animals ‘rarely take the course of abandoning these groups of humans by lumping them in with other animals.’³¹

Singer continues to cite William Frankena’s *The Concept of Social Justice*, in which the latter states:

...all men are to be treated as equals, not because they are equal, in any respect, but simply because they are human. They are human because they have emotions and desires, and are able to think, and hence are capable of enjoying a good life in a sense in which other animals are not.³²

Supporting my own observations discussed later in relation to Antonio Damasio’s Somatic Marker Hypothesis, Singer notes that Frankena’s definition of a human is reliant upon factors that are observable in non-human species; emotions, desires, the ability to think and to enjoy a ‘good life’, which Frankena explains is simply a ‘happy or satisfactory life’, rather than one with any specific moral underpinning. Given that definition, Frankena fails to demonstrate any solely human qualities and ignores morality, which may arguably have been a distinctly human trait, even if only through our inability to effectively measure it in other non-human species. Accordingly, Frankena’s definition is not one that I will directly apply to the key texts in this thesis, though

²⁹ Relph, S., and Foster, P., 2007, *Twin Girl With Eight Limbs to Have Surgery*, in The Telegraph online, in <http://www.telegraph.co.uk/news/worldnews/1568406/Twin-girl-with-eight-limbs-to-have-surgery.html>, accessed 5th January 2011.

³⁰ Skinner, M., in http://hrsbstaff.ednet.ns.ca/mskinner/Sociology/Socialization/feral_children.htm, accessed January 5th 2011.

³¹ Singer, P., *All Animals Are Equal*, in Regan, T., and Singer, P., (Eds.), 1989, *Animal Rights and Human Obligations*, New Jersey, p.8, in <http://www.animal-rights-library.com/texts-m/singer02.htm>, accessed 24th February 2011.

³² Ibid, p.7.

Singer's conclusions on the former's work do offer support to the possibility of the representations of robots and cyborgs in my selected texts being deemed persons:

Surely every sentient being is capable of leading a life that is happier or less miserable than some alternative life, and hence has a claim to be taken into account. In this respect, the distinction between humans and nonhumans is not a sharp division, but rather a continuum along which we move gradually, and with overlaps between the species, from simple capacities for enjoyment and satisfaction, or pain and suffering, to more complex ones.³³

I find this notion of a continuum an interesting one, and an idea that, when viewed concurrently with scientific and technological change, may help to explain the erosion of the binary oppositions between humans and robots in science fiction texts. The overlap of species referred to by Singer then, in the context of this thesis, could refer specifically to cyborgs, or more generally to the 'blurred line' that I referred to earlier, where it is difficult for members of the audience to distinguish between human and non-human in my key texts. Singer's discussion of simple and complex capacities suggests that it is humans who are capable of experiencing the latter, and therefore of being able to experience more emotions and sensations (both qualitatively and quantitatively) than other non-human species. This leads me at this stage to suggest that the human brain, bigger than any other animal's relative to body size,³⁴ is the ultimate differentiating factor given its role in consciousness and emotions.

Locke's concept of a person suggests that consciousness is the key not to humanness, but to personhood. In contradiction to Cartesian dualism, Locke maintains that the root of one's identity over time is in their awareness of their own emotions, physical body and place in time. Summarised by Peter Gibson:

1. A person cannot be a substance, because we never experience the substance; the underlying substance is irrelevant.
2. Consciousness is the experience which creates personal identity.
3. A 'man' is a creature (whose identity consists, like a tree, of its life), but a 'person' is a particular type of consciousness.
4. A person has the properties of being conscious, self-aware and rational.
5. Consciousness covers current mental states, awareness of our own bodies, and awareness of the past.
6. 'Person' is a forensic term, involving praise and blame, and a capacity to obey laws.³⁵

A clear distinction is made in the third point between humanness and personhood, which suggests again in conjunction with points four and two, that the representations of robots and cyborgs focussed upon in this thesis could be considered persons with their own artificial identity

³³ Ibid.

³⁴ Iredale, M., 2008, *Sci-Phi: Putting Descartes before the horse*, in *The Philosopher's Magazine*, Issue 42, 3rd Quarter 2008, p.41.

³⁵ Gibson, P., *John Locke's Theory of Personal Identity*, in <http://www.philosophyideas.com/files/mind/Locke%20on%20Personal%20Identity.pdf>, accessed 24th Feb 2011.

given that programmed rationality is a key feature in all of the robot/cyborg protagonists, as is self-awareness and awareness of others as individuals. Each one of those robotic characters also at least describes the mental states noted in point five (see later section on Damasio's Somatic Marker Hypothesis for further discussion of this issue), alongside an awareness both of their own bodies – Sonny's alloy; Andrew's inorganic parts, for example – and of the past, including Marcus's past human life.

The thoughts of a number of theorists who have criticised Locke's ideas are also noted in Gibson's text. The most relevant to this thesis is Thomas Reid's (1785) objection that 'We judge the identity of other people by their bodies, not by their consciousnesses'. This may well explain why the human resistance members in *Terminator Salvation* (and perhaps members of the audience too) were so willing to accept Marcus respectively as humans, compared to the rejection and scrutiny suffered by Sonny and Andrew.

Consciousness as a criterion for personhood, though, is noted in the work of Thomas Metzinger, Robert W. Mitchell and Daniel C. Dennett. The former identifies 'globality' and 'presentationality' as constraints on a 'minimal concept of consciousness'.³⁶ Globality places conscious states as part of an integrated world model. In terms of the robots and cyborgs at the heart of this thesis, each does interact with the world around them and recognises their place within it, though an amount of their 'understanding' of it is programmed rather than learned. That awareness, though, supports some of Locke's criteria for consciousness.

Presentationality, Metzinger continues, is a 'temporal internality – an island of presence in the continuous flow of physical time.' He cites Ruhnau's (1995) explanation, '...whatever I experience, I experience it now.' There is a clear link to the importance of time in this particular constraint, and thus a link to Prigogine's notion of science seeking to eliminate time (discussed in 'Rejecting Free Will'), and Bill McKibben's stance against transhumanism. Sonny, Andrew and Marcus all show a clear awareness of their future and past, creating a representation of a greater level of consciousness, and therefore a stronger suggestion of personhood.

The work of Mitchell and Dennett provides some more specific criteria for personhood that I will apply in this thesis to the study of the representations of robots and cyborgs in the key texts in order to ascertain whether or not they may be deemed persons, and therefore what the difference is between a person, and a human person. Mitchell suggests that prerequisite characteristics of personhood include 'verbal communication, self-reflection and knowledge that others are

³⁶ Metzinger, T., 2003, *Being No One: The Self-Model Theory of Subjectivity*, in précis of <http://www.philosophie.uni-mainz.de/metzinger/publikations/precis.pdf>, accessed 21st Feb 2011.

persons.³⁷ In order to self-reflect and recognise personhood in others, an individual must be self-conscious; aware of themselves both as an individual, and a part of any relevant social structure.

Certainly the robot and cyborg protagonists in the key texts herein do meet these criteria. Sonny, Andrew and Marcus all communicate verbally and non-verbally, for example Sonny damaging the table in the interrogation room, Andrew's frequent gestures and the fascination with winking that he shares with Sonny. They all exhibit a degree of self-reflection, focusing on and evaluating their own behaviour and the consequences, and each one shows evidence of recognising personhood in others, be it Sonny's discontent regarding the impact his actions may have had on the life and death of the man he called 'Father', or Marcus's references to his former human life compared to the one he lives now as a cyborg, in relation to John Connor.

Based on a straightforward application of Mitchell's criteria, it appears that personhood, if not humanness can be attributed to the robot and cyborg protagonists in the three texts I have selected, though importantly not to all of their kind in each text – NS-5 robots other than Sonny, or the 'skin jobs' created by Skynet for example. These are the villains of their respective texts and as such they are not granted any real semblance of personhood in order that they are clearly distinguishable from the human protagonists who represent the audience. Therefore the viewer is able to explore their 'wishes and nightmares' in a controlled filmic arena with protagonist and antagonist clearly marked.

However, these criteria seem rather simple on the face of things (though the mechanisms by which each is achieved are less so), as they omit emotions and other inter-related thought processes (according to Damasio, at least) such as rationality and purpose. Dennett, then, offers a more developed set of criteria in *Conditions of Personhood* (1976), cited in Mitchell's text. Six specific criteria are put forward, the first four of which Dennett asserts are exhibited by most intelligent beings, and thus it is the final two criteria that ultimately determine personhood. He states:

...personhood derives first from three mutually interdependent characteristics: being rational, being intentional and being perceived as rational and intentional. Once a being is acknowledged to have these three characteristics, personhood requires that the being reciprocate by perceiving others as rational and intentional; next the being must be capable of verbal communication and finally of self-consciousness. The last three characteristics are hierarchically dependent, building upon the first three.

³⁷ Mitchell, R.W., 1993, *Humans, Nonhumans and Personhood*, in <http://www.animal-rights-library.com/texts-m/mitchell01.htm>, accessed 19th November 2010.

Awareness of all but the last criteria is not a requirement for personhood. The problem for studies like Mitchell's is in determining whether or not the species being studied is capable of verbal communication, and whether or not they are self-conscious, particularly in the case of some animal species that we know little about and may therefore struggle to interpret their behaviours and social structures. However, as this thesis is based upon the representations of robots and cyborgs which communicate in spoken English, this problem is non-existent, and each of the film protagonists clearly satisfy Dennett's fifth criteria.

Mitchell notes that 'by self-consciousness, Dennett means that one is capable of reflective self-evaluation, that is, of adopting toward *oneself* the stance not simply of communicator but of ... reason-asker and persuader'. Again, there is clear evidence that each robot or cyborg in the key texts can reflectively self-evaluate, and therefore further support is given to labelling them persons. Sonny retrospectively studies any role he may have had in Lanning's death. Andrew traverses the entire narrative on a journey of self-evaluation and self-improvement (in his view, at least), as he replaces his mechanical parts with others more representative of human versions, and reflects upon the outcome in terms of his movement towards acceptance as a human person, not just a person. Marcus ponders his human life before his execution, particularly after the revelation that he exists as a cyborg after his revival. It is important to remember, though, that I am applying Dennett's criteria to determine personhood, not humanness.

CHAPTER ONE

I, ROBOT: TECHNO-UTOPIANISM

The aim of this chapter is to ascertain the degree to which Sonny may or may not be considered human, and consequently to identify the differences between a person and a human person that may constitute Dennett's 'special ingredient'. This will be achieved through the application of a range of theories, including those of Dennett and Damasio, to an analysis of the central characters and to a lesser degree, the mise-en-scene of *I, Robot*. Though the central focus of this chapter is Proyas's text, references to *Bicentennial Man* and *Terminator Salvation* will be made where they exemplify a similar point.

I, Robot is a relatively superficial exploration not just of the binary opposition between human and robot upon which this thesis is built, but of humanness itself, and the identification therefore of Dennett's 'special ingredient'. After all, in a futuristic world delicately balanced by programming and laws designed to protect 'us' from 'them' and allow humans and robots to co-exist in a hierarchical structure from which we directly benefit, knowing which group you belong to is clearly important. As such one of the questions we may wish to ask, though the film does not, is how can a robot identify a human? Given the range of possible criteria for humanness, this question is clearly a problematic one. Dr. Sam Vaknin describes the definition of 'human' as circular, a 'definition by negation: that which separates us from animal and machine is our human-ness'.³⁸

However, attempts to quantify this humanness often run into the vague or philosophical, rather than fixed, biologically-based definitions. Behavioural unpredictability and humour, discussed earlier, are traits frequently offered as resolutions to the debate, and certainly are traits present in each of the texts studied herein which contribute to the humanness of the central characters, but do not individually define it. In *I, Robot*, Spooner, Sonny and Calvin all demonstrate different levels of predictability in their behaviour, and at least initially it is the robot who acts more rationally, whilst Spooner's often foolish behaviour is foregrounded. Similarly, Spooner's jokes fall on unresponsive ears when directed at the aesthetically and emotionally robotic Calvin, more accustomed to reason and programming than spontaneity. Nonetheless, there are clearly a greater range of biological and emotional differences between the robot and its creator and pursuer than irrationality and timing.

³⁸ Vaknin, S., *On Being Human*, in <http://samvak.tripod.com/human.html>, accessed 8th June 2010.

Humanity as a social construct or product of biological ancestry are notions presented by constructivists and sociobiologists which again offer some explanation and are represented in my chosen texts, but are still not a resolution in themselves as they leave too many questions unanswered. For the filmic representations of these robots to fully interact with their human masters, a working knowledge of social norms relevant to their geographical and socio-economic working environment is required, and usually programmed in though hindered by our inability as creators to foresee and account for every possible scenario a robot may encounter.

Locke's theory of mind and the concept of a tabula rasa – that our personal identity is shaped predominantly by our nurture, rather than our nature and is therefore diachronic – is relevant here, as therefore is the ability or otherwise of robots to augment and update their programming to include these new experiences and thus develop an 'artificial identity' even if that is, as Vaknin queries, possibly an oxymoron. Time here is raised as an issue by Locke as it was earlier by Metzinger. Further, it is also a concept both addressed in the key texts studied herein, and potentially relevant to a criterion for humanness.

Morality and empathy are issues which link with both Locke's theory of mind and Dennett's criteria for personhood – the ability to recognise other humans because we too are human, is a part of our humanness. Sonny regularly identifies himself as different to 'the other robots', and thus demonstrates self-evaluation. As an emotional trait, empathy goes beyond the corporeal (though the events leading to the empathic state may have resulted from physical interaction) and begins to move the debate towards the aforementioned theories of Descartes.

DETECTIVE SPOONER

Before the opening scenes of *I, Robot* even begin, images on the DVD menu screen depict the technicalities of a robotic 'brain', the pistons and rods that connect the head of an NS-5 to its body, and the mechanics of the robot's arm, notably the key factor shared by Spooner and his robotic counterpart. Alongside repeat images of Sonny's eyes – described by humans as the windows to the soul – these are the 'lights and clockwork' that Spooner later refers to in the film when damning robots for their lack of empathy.

The early focus on the brain appears to lend some support to the view that the brain is the key to humanness, and to the erosion of the binary opposition between robots and humans in contemporary Anglo-American Science Fiction films such as this. Of course, if the eyes were literally a window, it is the brain that one would see when looking through them, rather than the

implied Cartesian separateness of something far more incorporeal. Sonny's eyes are blue, notably, further distinguishing him from others of his kind.

Images of Sonny are juxtaposed with those of Spooner, whose facial expressions convey negative connotations to the viewer of his 'proximity' to Sonny. Additionally an image taken from Sonny's interrogation scene is used which shows the robot and the Detective at opposite ends of the table over which Sonny's lack of humanness is debated. It would appear, therefore, that the Detective and his quarry are being presented as opposites in more ways than simply their role. The organic versus inorganic debate is represented by these images at the film's outset, and set the tone for the audience's expectations of Spooner, who represents them, and his actions towards anything non-human.

Shortly after entering the United States Robotics (hereafter referred to as USR) building to begin his investigation into the death of Alfred Lanning, Spooner encounters Dr. Susan Calvin, another robotics expert charged with making robots "seem more human", who herself has an ambiguous past with Lanning and who rejects any suggestion by Spooner that a robot could defy Asimov's three laws of robotics (though here they are attributed to Lanning), and harm its creator. It is worth noting that his investigation and attitude are arguably already biased as, if personal identity and beliefs are based on experiences – nurture – Spooner's perception of himself and others and his interaction with them, specifically Calvin and Sonny here, will be altered accordingly.

This in part offers an explanation of his suspicion not just of Sonny, but of the three laws of robotics designed to protect humans from robot harm. When these idealised laws are studied in detail, it would appear there is cause for Spooner's concern arising from the non-human nature of the robots which are charged with upholding these laws and their resulting lack of ability to apply them consistently. For example, each of the key texts has a significant focus on the human brain and heart as key to differentiating humanity from robots. Yet the first law makes no specific reference to the prevention of any emotional or mental injury to humans, despite the link that would be drawn between said damage and these privileged organs. How could a robot determine what may cause mental or emotional harm to a human, and decide on an appropriate course of action to prevent it, when they are limited by programming?

It would take the kind of understanding of human society and bounded rationality proposed by Locke to determine at what point the said emotional or mental damage becomes serious enough to warrant action. Imbuing filmic representations of robots with this capacity therefore somewhat 'humanises' them through the laws that were designed to uphold the differentiation between 'them' and 'us'. It is Spooner's apparent awareness of their incapability in this sense that triggers

much of his suspicion, as when viewed in this sense, robots programmed with the three laws may be more dangerous than those without, as the human programmers behind such creations are themselves incapable of foreseeing every possible circumstance for which a robot would then be programmed accordingly.

How could we (or in this case, Spooner) identify a rebellious robot from one that is malfunctioning or simply lazy, as we are not privileged to the point of transition between the two, and have no way of measuring the change or indeed attributing blame thereafter? It is this inability to distinguish intent and the potential for robots to override their programming that identifies films like these as dystopian scenarios. In Spooner's case, his suspicions are increased because he too has no way of determining Sonny's intent, and without a concrete answer to the death of Lanning sees only a dangerous rebel in Sonny, continually defended by Calvin, who aesthetically appears to be more robot than human herself.

As the actions of the key protagonists in *I, Robot* are largely based on a plan devised by Lanning and triggered by his death, we make another return to the issue of free will and determinism, and I wonder if Susan Calvin was so called by Asimov in the original short stories upon which this film was based, in reference to Calvinism – John Calvin specifically – and the determinist nature of that school of thought. Whether or not she was, I personally subscribe to the idea of bounded rationality, discussed further later. Though humans, and indeed non-human animals and robots or cyborgs – be they real or filmic representations – have the physical capacity to act 'freely', learnt rules of what is and is not socially acceptable, the existence of a conscience (at least in humans) may often prevent us from acting in a specific manner despite our ability and perhaps desire to do so.

A dog is likely to learn quickly that acting on its desire to dig up a garden will result in negative reinforcement. A robot may be physically capable of performing acts outside of its programming, but is nonetheless constrained by it. A human may wish to commit a crime and indeed be capable of doing so, but refrain based on the potential consequences, either to themselves or to others, determined by their conscience or need for self-preservation. Our central protagonist is an officer of the law, so his bounded rationality is overt. Susan Calvin, however, is bound not only by the same laws that Spooner upholds, but by her own personal and professional beliefs, standards and experiences – right down to her reserved visual appearance.

DR. SUSAN CALVIN

A complex character, Susan Calvin's humanity is at no point questioned by the narrative (or audience), be that because or in spite of her function as something of an 'alter-ego' for Spooner. Though Calvin gradually moves away from convictions based purely on logic throughout the course of the narrative, when initially introduced to the audience she appears both visually and behaviourally robotic. Apparently organic, she displays feelings of sadness for Lanning alongside anger, frustration and bemusement at Spooner's unwillingness to accept the apparent technoutopia that surrounds him. Yet both her extreme rationality and visual appearance connote something altogether less human.

Dressed in a long silver suit, hair fixed back perfectly, and rarely seen comfortable outside of a laboratory, Calvin looks not dissimilar to one of her NS-5s. Her professional stance and rigid attitude alongside her frequent close proximity to, and defence of Sonny, distance her from Spooner as she snubs his humorous attitude and casual observations, preferring instead to rely on rules and laws, much as her robots do. Humour, it is noted in *Bicentennial Man*, is an inherently human trait, albeit not an outright definition of humanness. Further, as Spooner acts as the audience's guide to a representation of their future - linked to them by his Frankenstein complex, sneakers and their star expectations of Smith as the minority underdog so often charged with saving the world they inhabit - Calvin's behaviour also distances her from the audience. As such the principal component of humanness presented in *I, Robot* appears to be emotion as opposed to rationality.

Yet on the other hand there is evidence of Calvin's vulnerability and emotion, traits which move to the fore as the narrative progresses. Typically this occurs away from the comfort zone that is her laboratory, and when confronted with a situation in which her logical mind has no option but to accept Spooner's explanations, for they then seem the most rational - her restraint by a VIKI-controlled NS5 for example. She, like her robots, functions effectively and calmly when in a controlled surrounding with familiar mise-en-scene such as her laboratory. However, once variables such as representations of individuals with free will are introduced in a social setting, Calvin's reliance on a priori judgements and unwillingness to consider Spooner's alternative explanations for events bring chaos to her mental order and poise.

This is not to suggest of course that Calvin becomes in any way frantic or uncontrolled. Far from it. But Calvin's work life and arguably social life, as we are shown that she spent a part of that time with Lanning, have been spent operating within Herbert Simon's bounded rationality, and creating such rationality for others through her programming. USR, for whom Calvin works, has a

massive presence and importance across *I, Robot's* representation of America and in the homes of the vast majority of its inhabitants. Visually, the building towers above its surroundings and stands in stark aesthetic contrast to the preferred haunts of audience-favourite Spooner, and the deep, dark watery depths that continue to haunt him, stretching toward the sky in the words of Richard Curran Trussell 'like the tower of Babel', representing 'the hubris of humanity to attain godlike power by ingenuity and guile'.³⁹

Such emotional constraints and constant focus on reason and logic present Calvin with situations such as her aforementioned control by one of her creations where she encounters what Lanning described as 'free radicals' which '...engender questions of free will, creativity.' Such circumstances do produce emotional responses in her, resulting in Prigogine's 'unstable order' and supporting Damasio's assertion that emotion and reason are in fact not separate, but interdependent. Similarly, as a detective, however driven by resentment and haunting memories, Spooner must also exercise reason alongside emotion. Trussell offers some support to the idea of linked emotion and reason:

Both intellect and emotion define us, yet congruent with ancient peoples, the heart is the true seat of consciousness and conscience, the root of emotion, and the link with something mysterious that cannot be captured or domesticated by the intellect.⁴⁰

Whilst the human heart is frequently represented both in films such as those studied here, and socio-cultural events (e.g. 'giving your heart' to someone on Valentine's day) as Trussell's 'root of emotion', frequent references to the human brain in the key texts suggest that there is more than one 'true seat'. The human brain is a massively complex organ and, unlike the heart, cannot be accurately replicated in function by science and technology (or in Trussell's words, 'the intellect'). Though Trussell does not state that in the organic versus inorganic debate that it is the heart which defines human beings, the level of importance he places solely on the heart is to the exclusion of the brain. In the filmic examples studied herein, there are frequent references not only to the heart, but to the importance of the brain, including in the film which Trussell is critiquing.

Damasio, too, argues that emotion and reason are not exclusive, but rather are often interdependent with the ratio of one to the other being controlled by the experiences, culture, beliefs and so forth of the individual, explaining the differences between Calvin and Spooner's approaches. Damasio's views though, which I support, unlike Trussell's, place the emphasis on the brain. Those experiences I refer to are interpreted via the sensory receptors on the individual

³⁹ Curran Trussell, R., *I, Robot: You Gotta Have Heart* in *The Journal of Religion and Film*, October 2005, Vol 9., <http://www.unomaha.edu/jrf/Vol9No2/TrussellIIRobot.htm> , accessed 3rd August 2010.

⁴⁰ Ibid.

concerned – a human’s five senses in our case, and that of the representations of humans in the key texts here – therefore as a robot like Sonny may have similar sensors, we may be open to the possibility that he is actually experiencing some of the emotions he describes.

Therefore, if Sonny and the other automata are aware of their own mental states and bodies as a result of their sensors, they are satisfying certain criteria for personhood, such as the fifth point in Gibson’s earlier summary of Locke, as those are elements of the definition given of consciousness, and consciousness is noted as a criterion of personhood by Locke, Dennett and others.

Two problems then are raised, the first regarding the capabilities of said robots and their sensors. This thesis has thus far asked what it is that makes a human, human. Given the writer’s organic nature, no real question over whether or not humans could recognise that ‘special ingredient’ has been asked. Though it is not specifically the focus of this thesis, it is worth noting that a lack of ability in one human to recognise another would in turn impact on the ability of a robot to identify a human, as they are programmed by us. Remembering that Robert Mitchell noted a ‘...triple reflection of consciousness, [...] that I recognise that you recognise that I have consciousness’⁴¹ as necessary for personhood, based upon that assertion such a lack of recognition on the part of a robot would negate its ability to be considered a person.

So given that the reasoning and sensory capacity of robots is determined by their parts and encoding, the fact that this thesis is searching for the answer to the question of what it is that distinguishes person from human person, long after the creation of many robots, leads me to wonder what answer to this question of distinction might be embedded into those robots’ circuitry? I noted Hari’s comments earlier regarding the inability of robots to correctly identify different types of fruit. Perhaps in the same way as *Bicentennial Man* suggests a focus on dying, rather than living as a human, when endeavouring to define ‘person’ and ‘human person’, we should consider not just how a robot distinguishes itself from a human, but how a robot discriminates between a human and other organic species. Given that each of the robot or cyborg filmic protagonists serves humans in one way or another, this recognition is especially important with regard to the compliance of the former with laws like Lanning’s that govern their behaviour.

The heart, of course, is an organ common to many species, thus its presence does not determine whether an individual is human or not. The same is true of the brain. Whilst its absence is enough to determine that an individual is not human, its presence does not guarantee the opposite.

⁴¹ Mitchell, R.W., 1993, *Humans, Nonhumans and Personhood*, in <http://www.animal-rights-library.com/texts-m/mitchell01.htm>, accessed 19th November 2010.

According to a gene study⁴² reported in National Geographic News in 2005, humans and chimpanzees are genetically 96% identical. In fact, according to primate scientist Frans de Waal 'We are apes in every way, from our long arms and tailless bodies to our habits and temperament.' The report continues 'The number of genetic differences between humans and chimps is ten times smaller than that between mice and rats.' How then does a robot distinguish a human from a chimp by a simple scan and by extension accurately follow the Three Laws? The narrative of the *Terminator* franchise would have been entirely different if Skynet's creations could not differentiate between human Resistance members and their own kind.

The second problem raised can be exemplified using Sonny's interrogation scene in *I, Robot*, when Spooner's questioning about his ability to 'simulate emotions' results in Sonny hitting the metal desk at which he is sitting so strongly that imprints of his fists are left behind. 'That one's called anger. Ever simulate anger before?' taunts Spooner. His interrogation of Sonny is hinged around the detective's belief (one likely shared by the majority of the audience) that robots are nothing more than 'lights and clockwork' with no capacity to think for themselves or to experience emotions. Again, such descriptions are references to the brain and heart.

A range of emotions and behaviours are raised throughout the narrative as human behaviours, but just as possessing a heart, brain and walking upright doesn't define a human, love, hunger, sleep, dreams, trust and creativity are all traits exhibited by humans, but not exclusively. Thus the debate between Spooner and Sonny in the interrogation room is little more than an exchange on the organic versus inorganic issue; what or who is alive or not, rather than what it is to be human.

Interestingly, this is one of the key scenes in the film that begins to mark a shift in the personalities of three of the key characters (though Sonny's behaviour is not necessarily a personality, of course). Sonny and the arguably robot-like Calvin appear to be exhibiting emotions as the more animated Spooner follows Lanning's trail of breadcrumbs. If Damasio's thinking is to be applied here, specifically to Sonny, it is possible that Spooner's quip about simulation is noteworthy if Sonny is simulating these 'emotions' rather than actually feeling them. That simulation may simply be an innocent part of Sonny's programming given Calvin's job is to 'make the robots appear more human', which it is not stated is limited solely to aesthetics. Conversely it could be more breadcrumbs or misleading for Spooner.

⁴² Lovgren, S., 2005, *Chimps, Humans 96 Percent the Same, Gene Study Finds*, National Geographic News, p.1, in http://news.nationalgeographic.com/news/2005/08/0831_050831_chimp_genes.html, accessed 5th August 2010.

Returning to the earlier discussion based on the bi-directional nature of interactionism and a human's contact with their surrounding environment, Damasio argues that mental events both precede and follow physical events, one as a consequence of the other. Lanning knew Detective Spooner would follow his Hansel and Gretel-like trail not simply because of his personal familiarity with Spooner and knowledge of his aptitude as a law enforcement officer, but because of his haunting past. The mental events that robots evoke in Smith's character – memories of rescue from slowly drowning whilst watching a young girl die instead; the icy temperature of the water; the pain in his damaged arm; his helplessness to change events – are based on his interaction with his environment through his senses.

This interaction generated mental events, originating in the brain, which in turn led both then and now to physical events, for example trying to break away from his robotic rescuer to save the girl and chasing down a robot that was innocently taking medication to its owner based on a suspicion that it was committing a bag theft. Though Spooner's heart would undoubtedly have responded physically to both sequences of events, and he clearly maintained a sense of morality throughout, however biased, as discussed earlier physical events are both created and the resulting chemical messages received in the brain, not the heart. Thus even though both organs may generate or respond to an emotion, the mental event and following message for the body to act originate in the brain. Therefore it is my belief that greater importance should be placed on the brain in filmic representations such as these.

However, human beings have the capacity to be simultaneously rational and emotional creatures. Given that both Spooner and Calvin exhibit emotion and reason to varying degrees, these characters do not represent us, the audience, and them, the robots respectively; instead they each represent a side of the human condition identified by Damasio – not distinct, but interdependent, just as Calvin and Spooner come to be.

We do not question whether or not the emotions displayed by Spooner and Calvin are genuine in the same manner as we do with Sonny, perhaps because we can identify with characters we know are representations of beings that are born, not manufactured, though that in itself, of course, does not make them human. As such, they possess a human brain and therefore have the capacity to fully experience emotions according to Damasio's Somatic Marker Hypothesis. Further, because we are shown the root of Spooner's traumatic memories and are given a brief insight into distressing times in Calvin's life, through empathy arising from shared experiences or sympathy from an understanding of the potential mental and physical impact of such events that may have occurred in our own lives, we as humans understand their feelings and believe them to

be real, despite our knowledge that what we are seeing is a representation, and thus all feelings that arise from the text are a mechanical reproduction. As empathy and sympathy are both emotional states and linked to a conscience, they too according to Damasio are reliant on the human brain.

SONNY

Sonny is an NS-5 robot created by Lanning who refers to the doctor as 'Father'. As *USR*'s latest output, he is at the pinnacle of advancement of robot servitude. But Sonny – clearly named to highlight the issue of birth versus construction, organic versus inorganic – has been created differently from all other NS-5 robots, with denser alloy and, according to the robot, the ability to dream. He is Lanning's 'ghost in the machine', a tool designed to eradicate extreme robotic control and restore the hierarchy between man and machine. In an interview post-release of the film, Smith argues that "the robots aren't the problem. The technology is not the problem. It's the limits of human logic that is the problem, and we are our own worst enemy."⁴³

Linking to Trussell's work on the religious imagery in the film, the robot's name then also forms something of a biblical reference for Christians, as the Son of the Father whose role it was to restore balance after we became 'our own worst enemy'. Religion requires faith, often viewed as an exclusively human trait, and also a level of self-consciousness outlined by Dennett⁴⁴ as one of the requirements for personhood.

It is Sonny's difference in design from all other robots of his kind that afford him this ability. By virtue of the fact that he is a creation, built not born, his sculpting links him to Prigogine's notion of art and its temporal quality. Sonny's design not only demarcates him as a robot – one of a breed of new and improved robots who eventually rebel against their creators – but is such that he will last forever. Immortality is seen in *Bicentennial Man* as a reason for rejection by humans, in *Terminator Salvation* as a quality possessed only by the enemy and highlighted by Marcus's sacrifice, and thus in *I, Robot* as a fact upon which the audience and Spooner may base their suspicions, especially as Spooner's inherent dislike of robots is rooted in the mortality of another.

⁴³ Smith, W., cited in Richardson, K., in <http://www.culturewars.org.uk/2004-02/irobot.htm> , accessed 7th April 2010.

⁴⁴ Dennett, D.C., cited in Mitchell, R.W., *Humans, Nonhumans and Personhood*, in Cavalieri, P., and Singer, P. (eds.), *The Great Ape Project*, 1993, p.237-247, in <http://www.animal-rights-library.com/texts-m/mitchell01.htm>, accessed 19th November 2010.

In the opening minutes of *I, Robot*, we are witness to the memories that haunt Spooner in his sleep surrounding the death of Sarah, a young girl that a robot calculated had less chance of survival than the Detective it subsequently saved. As a result, Spooner lives with a constant feeling of guilt, borne out throughout the film in his attitude towards USRs creations, and in comments to his Grandma when she wants a new robot that 'no good ever comes from them'. Why should he have lived when a young girl died? Far from the dissimilarity between Spooner and Sonny that the DVD menu implied, both characters live with blood on their hands.

Smith's character struggles with memories of the tragic death of a child for which he feels responsible. Sonny simulates guilt and uncertainty over Lanning's death – a death he wonders if he could have prevented, and which occurred so that Sonny could ultimately stop VIKI. I have specified that Sonny simulates rather than feels guilt, as to suggest it is a genuine emotion would be to imbue Sonny with a level of emotion and associated humanness that I don't believe he has; it is Sonny's alloy that is modified, not his 'brain'. However, from the audience's point of view both characters are similar in more than their possession of robotic arms, and grow to become more like each other throughout the narrative, mirroring the trajectory of the binary opposition between human and robot in more recent Science Fiction films.

Spooner argues throughout for individuality and freedom of choice being human qualities, qualities which at least in part, Sonny displays. Given the high regard in which humans hold their individuality, and the presentation of the banks of NS-5s as homogenous artefacts, this difference in Sonny adds to the audience's acceptance of possible personhood. Yet it is hard to ignore the suspect's visual appearance, and he is subject to the fear of 'the other', the 'Frankenstein complex' through Spooner, often raised by Science Fiction texts as one of the more negative traits of human identity. From a discussion between Spooner and Lt. John Bergin:

Bergin: I've been thinking. This thing's like the Wolfman.

Spooner: Uh-oh, I'm really scared, John.

Bergin: No, seriously. Guy creates monster. Monster kills guy. Everyone kills monster.
Wolfman.

Spooner: That's Frankenstein.

Spooner of course is technologically augmented with an inorganic arm after the aforementioned car accident, which he is reticent to reveal, as his dislike of the inorganic extends to part of his own body. However, there is still no question raised regarding his humanness despite his prosthetic. Like Andrew in *Bicentennial Man* he discusses the corporeal differences between himself and robots as originating in the brain and heart: he notes that VIKI has 'a brain' and that robots (motioning to his chest) '...have nothing here, just lights and clockwork. Go ahead, you

trust 'em if you want to.' Importantly, trust he claims is '...a human thing' that Sonny wouldn't understand.

There is some support given to Spooner's emotional outburst by Damasio, who states that 'the absence of emotion and feeling is no less damaging' than the '...potentially harmful influence of emotional biases.'⁴⁵ Oddly though, it is perhaps Spooner who exhibits the least trust of any character in the film, and Damasio continues that this damage may be done to 'the rationality that makes us distinctly human', the same rationality that Spooner berates Sonny for.

As Dennett claims that personhood '...derives from three mutually interdependent characteristics: being rational, being intentional, and being perceived as rational and intentional'⁴⁶, Sonny's behaviour throughout the film – be it his apparent awareness of his difference from other NS-5s, of his own decommissioning and his resulting 'fear', or his intent to stop VIKI – does demonstrate rationality and intent, and therefore support the case for considering him a person.

Through his refusal to become any more a part of the technological society in which he lives than surgery has forced him to be, Spooner embodies what transhumanists would describe as a 'humanish'⁴⁷ attitude, seen recently in *Surrogates* (2009, Dir. Johnathan Mostow) in the Human Coalition, and less specifically (with regard to the fact that there was no required alteration to the human form, with the exception of Marcus Wright) in films such as the *Terminator* franchise, built upon the human resistance to robotic take-over. However, Spooner's reasons for this rejection of the human-enhancement technologies endorsed by transhumanists are not based on moral objections to the alteration of the fundamental aspects and limitations of the human body, but upon damaging personal experiences that have led him instead to fixate on the limitations of robots.

This obsession is demonstrated throughout the film, for example in his discussion with Sonny regarding whether or not a robot can turn a blank canvas into a piece of art, or write a beautiful symphony, and his angry rebuffing of robot's lack of emotional intelligence after one saves him from a car accident rather than a young girl, who subsequently dies, based on a mathematical probability. As such, Spooner's level of emotion is far higher than that of other characters, human and robot alike, and the audience consider him human, unlike the rational, apparently uncaring robots alternatives. Sonny's response to Spooner's questioning about his artistic abilities though

⁴⁵ Damasio, A., 2006, *Descartes' Error*, Introduction, p.xxii.

⁴⁶ Dennett, D.C., cited in Mitchell, R.W., 1993, *Humans, Nonhumans and Personhood*, in <http://www.animal-rights-library.com/texts-m/mitchell101.htm>, accessed 19th Nov 2010.

⁴⁷ Alexander, B., 2000, *Don't Die, Stay Pretty: Introducing the Ultrahuman Makeover*, in <http://www.wired.com/wired/archive/8.01/forever.html>, accessed 6th Dec 2009.

– “Can you?” – serves to highlight Spooner’s human limitations, and suggest that as Sonny is a human creation, perhaps the lack of emotion for which his kind are berated by Spooner, is in fact the result of human inability to create an effective programme.

Individuals like environmental ethicist Bill McKibben have long argued against the concept and practice of transhumanism given their view that it would be morally wrong, and would reduce our experience of life by removing the urgency of time and the additional meaning that adds to the choices we make.⁴⁸ Personally, though I understand McKibben’s view and partly agree that in some scenarios transhumanism might reduce our experiences, I believe it also opens up a range of other opportunities, particularly for disabled individuals. Therefore to me, the issue with transhumanism is in the reasons for the practice, not the outcomes.

McKibben’s argument though is similar to the assertions of Prigogine, discussed earlier, about the negative consequences of the elimination of time by science, and also to the human rejection of immortality displayed in *Bicentennial Man*. There it is argued, and will be discussed in greater detail later, that it is our ability to die that makes us human, and allows Andrew to finally be accepted as such, albeit after his death. Thus the organic again becomes an important issue in this debate as it distinguishes ‘us’ from ‘them’ – human from robot – given our ability to die, which in turn increases our emotional responses and highlights the importance of the heart and brain, privileged so clearly in *Bicentennial Man*, and referred to throughout *I, Robot*.

By our current standards, Spooner’s attitude would probably be classed as ‘retro’ and likely praised for its adherence to the recent fashion of reviving pop-culture icons that have, since their day, been only the preserve of the geek. Views like Spooner’s are of course archaic in the representation of his time that we are shown in the film, and his prejudices become the driving force behind his dogged pursuance of what Kathleen Richardson describes as ‘a mechanical whatdunnit’⁴⁹ - pursuance that not only confirms his belief in a robot conspiracy, but that reveals his true self to a somewhat misled audience, and provokes a re-evaluation of the criteria for humanity in both protagonist and viewer.

Of course, from the point of view of the audience still in that time of ‘retro appreciation’, this is a call for identification with the officer, a law-keeping citizen who should ideologically command our respect and loyalty. His fondness for what in 2035 would be obsolete technology and outdated fashion sense allows us entry to a futuristic world unfamiliar to us, and eyes through which to

⁴⁸ McKibben, B., 2004, *Enough: Staying Human in an Engineered Age*, Ch.1, *Too Much*, p.1-65.

⁴⁹ Richardson, K., in <http://www.culturewars.org.uk/2004-02/irobot.htm> , accessed 7th April 2010.

view it. Spooner's experiences then become ours, and he the reference point against which we judge new-age Chicago and its inhabitants. At the outset of the film we are deliberately kept unaware of Spooner's history and resulting status as a cyborg. Thus as a human, like us, his old-fashionedness, appreciation of what we have in our own time, and importantly his suspicion of anything different – primarily emotional responses – become beacons of humanity, even if they are not specifically addressed as such.

It would appear, then, that the obvious visual and hierarchical differences between creator and created in *I, Robot* are set to form the basis of an exploration of man versus machine - the organic contrasted against the inorganic. Not so, argues Richardson. Rather she believes that it is about 'emotion versus rationality, impulse and prejudice versus logic and reason.'⁵⁰ Whilst I agree with this statement in principle, her assertion that it is 'not a tale of man versus machine' is problematic, as those factors she accurately outlines as defining the narrative are typically presented in filmic texts such as this as the definition of and distinction between humans and robots, and their representations in Science Fiction films.

Granted, Richardson clearly states that it is not necessarily robots that are always represented as cold and rational, nor exclusively humans who are impulsive and emotional, but by Richardson's own admission, those who demonstrate emotion and impulsiveness 'are elevated above those who do not.' Spooner is the central protagonist, clearly represented in a manner which invites the audience to accept him as human (though he is in fact a cyborg) through his prejudice towards automata, and it is he who displays the most irrationality and 'human' tendencies. The closer Sonny becomes, both in personality and relationship to Spooner, the higher he is elevated in the esteem and value of the audience.

So whilst *I, Robot* offers the possibility of such traits being exhibited by an inorganic artefact, they are never truly separated from human ownership and typicality, thus the binary opposition between man and machine that Richardson rejects is still valid. The traits she lists are simply commonly offered definitions of man and machine. Further, if Damasio's argument is applied here, questions are raised over how genuinely Sonny is experiencing these events that Richardson refers to, when to be conscious, able to sense, does not necessarily result in emotional response.

However, although we may not grant him humanness, when judged according to Dennett's criteria for personhood, Sonny satisfies each condition. He is rational and intentional throughout, sometimes problematically so, for example in his determination to destroy VIKI as Calvin's life

⁵⁰ Ibid.

hangs in the balance – Sonny reasons that his actions may save many lives, not just one. Whether positively or negatively, Sonny is perceived as both rational and intentional, particularly by a suspicious Spooner in the early scenes of the film. He communicates verbally (though none of these films address the technological implications of this), and is aware of himself as an individual, and of his differences to others of his kind; blue eyes, denser alloy and so forth.

CHAPTER TWO

WHAT IS IT TO BE HUMAN?

This second chapter will draw on the key texts and other examples from film and television, outlining the perspectives from which they are analysed, and providing reasons for the rejection of alternative views. It will present a broader discussion of the ideas of Dennett and Damasio amongst others, and the centrality of the brain in these narratives, as the chapter explores the extent to which traits commonly deemed as human are shared by the filmic protagonists, and therefore the degree to which they may, or may not be considered persons.

In more contemporary films, the once clearly defined binary opposition between human being and robot has begun to erode, particularly so since Robby the Robot gave way to a new breed of part-human, part-robot cybernetic organisms – cyborgs. Representations of cyborgs, whether as visually obvious as the original *Terminator* and *Robocop* (1987, Dir. Paul Verhoeven) or as obscure as Det. Spooner from *I, Robot* literally blend human and machine to a level where it is now difficult to discern between the wholly organic human and partially organic robot.

Although the reasons for the ‘erosion’ may form some part of this thesis, its focus is instead on the level of erosion and the resulting personhood these cinematic representations of automata appear to have. Thus in order to address the question of ‘artificial identity’ we must ask what the difference is between human and robot that once constituted the aforementioned binary opposition. Does the answer lie in a distinction as simple as the organic versus the inorganic, or is the formation of the human self dependant on something greater than the sum of its parts?

The most obvious answer to this debate initially appears to be the issue of consciousness. It seems universally accepted, and this thesis will pre-suppose, that all humans to one degree or another are conscious beings. Therefore, conscious and wilful (though the nature of will is to be discussed later), humans each have their own unique identity shaped and developed according to individual circumstance - the nature/nurture debate - and informed by experiences. Certainly in the early days of the cinematic representation of robots it was clear from examples such as Fritz Lang’s aforementioned false Maria in *Metropolis* that had a soul transferred into it from its human model, that the robots represented in films were in no way conscious beings. Indeed, despite his exceptional intelligence and frequent attempts to enhance his programming with software that would augment his understanding of human rituals and experiences, even the android Data in *Star Trek: The Next Generation* could be quite simply deactivated. Therefore these creations,

however lifelike in appearance, had no level of humanity, only programming designed to create an increasingly human persona.

VIEWS AND APPROACHES

At this juncture it is important to identify the point of view from which the selected films are being addressed. Clearly there is a wide range of philosophical and scientific views applicable to this study of the erosion of the binary opposition between man and machine and the possibility of personhood in the latter. Thus clarity regarding which theories I am applying is of importance given the significant differences in reading that such varied theories may yield.

My approach is culturally determinist; determinist in the sense that I believe that events are causally determined and thus the state of an object – or specifically in the case of this thesis, the representation of a human, robot or cyborg – is largely determined by prior states and events. Accordingly, I do not believe in free will, thus a discussion of this is not an area of specific focus in this thesis. Further discussion of my reasons for rejecting free will is offered later.

Though numerous processes outside of the scope of this thesis (evolutionary processes for example) are argued by some theorists to be fundamentally indeterministic,⁵¹ as this thesis is dealing not with human evolution, but specifically with the representations of robots and cyborgs in filmic scenarios, my methodology is deterministic and materialistic. As such it is based upon the notion that matter is all that exists, and therefore is the cause of phenomena such as consciousness (often hailed as a differentiator between human and robot, albeit not exclusively human), via interactions that occur in the causal manner outlined above. Ideas then like Descartes' mind/body dichotomy are rejected in favour of more scientific explanations, thus the ideas of neurologist and neuroscientist Antonio Damasio are relevant.

I label myself specifically a cultural determinist in this study because I am endeavouring to establish what makes us human, and my belief is that our behaviours are primarily, though not singularly, the result of our culture and social environment as this determines who we are, and is an experience the robots are not privileged to. Cultural determinism represents the nurture side of the nature versus nurture debate.

Whilst it is my belief that the latter has the greater effect on the behaviours we exhibit, I do not believe this is to the exclusion of genetic make-up. An individual may be born for example, with

⁵¹ Shanahan, T., *The Evolutionary Indeterminism Thesis*, in *BioScience* (2003), Vol. 53, No. 2, p163 [in <http://www.jstor.org/pss/1314330>, accessed 29th October 2010.]

some impairment of the functioning of one or more areas of the brain, and thus be less able to develop in the related areas – socially, linguistically etc – that ultimately contribute to their identity, sense of self and personhood, and even their understanding of the bounded rationality that will be discussed shortly.

Conversely, robots do not have the same capacity to change and develop, to learn new laws and social rituals that humans do. They are programmed and thus limited by the programmer and the technology and finances available. The fact that robots do not socialise in groups as humans do – the nurture side – then becomes irrelevant, as they are generally not capable (in reality at least, though there are cinematic representations such as *Star Trek's* Data that break the rule) of augmenting their programming with any new information gleaned from such interaction in the same manner that a human would. In this sense a study based solely on robots would be more closely likened to biological determinism as their behaviours are fixed by their parts, as a human's would be by their genetic nature, assuming they were to forgo any transhumanist modifications.

I acknowledge the role that the human brain plays throughout this thesis in my analyses of the key texts, and thus make no suggestion that the 'nature' side of the debate be ignored, rather I would return to my criticism of Hayward, Cardinal and Jones. However, I would also argue that the focus here is on the presence or absence of a human brain and its irreplaceability by way of highlighting it as the ultimate differentiator between human and robot, rather than on the degree of functionality each representation of a human appears to have. Simply, I believe that there is a greater variation of socio-cultural backgrounds and experiences than genetic structures within the human race as a species (as well as in the protagonists in the texts studied herein), and therefore that the greater impact is likely to be caused by our socialisation and experiences than biologically determined factors.

The human brain then is the central 'hub' which interprets each experience – the causal interaction of matter referred to earlier – communicated to it by the body's sensory registers, and therefore establishes the next course of action in a causal, determinist manner that will be discussed in terms of Antonio Damasio's Somatic Marker Hypothesis. An area of focus in this thesis, then, is reference to the human brain in the key texts that suggest its inimitable nature, and thus its status as the factor that makes us human persons, rather than simply persons.

Further, my position is incompatibilist in that I do not believe in the existence of free will. I have stated that it is my belief that it is predominantly our culture that creates our identity and sense of self within our micro and macro sociological environments. As such a significant part of human culture internationally, regardless of otherwise differentiating factors between individuals and

cultures such as affluence or educational, medical, technological and scientific advancement, is the existence of laws, be they legally enforced or socially expected rules for living.

It is the knowledge of such laws and their actual or perceived implications that create what Herbert Simon describes as 'bounded rationality'⁵². Whilst a human may have both the opportunity and the physical and mental capacity to act in a certain manner – let us say, for example, to shoot someone – their knowledge of laws and consequences relating to that act will doubtless modify their ultimate course of action. Therefore, without wishing to enter into a debate about the extent to which morality and ethics are exclusively human traits, it is clear to me that life within this bounded rationality negates the possibility of truly free will through its restriction.

Rather I feel there is an illusion of free will that occurs because of the range of possible reactions to those finite laws from an equally extensive collection of individuals which makes the resultant behaviour unpredictable, not free, and still causally related to a previous event. In numerous Science Fiction texts, the human control of robotic 'free will' through programming or superior intellect (in *I, Robot*, Lanning's three laws and Spooner's detective skills) is a common theme used to explore Nichols' nightmare of us being usurped by our creations.

It is important to note that given the often complex variations in social rituals between different cultures, my discrimination between legal and social laws is not intended to imply that one is necessarily more complex or valuable than the other. Both may have an impact on the decision making process of an individual, human or not, and therefore the relative value an individual places upon either form is not a marker of humanness, but of perceived implication during a given event. That perceived implication will vary according to individual experience.

Whilst I believe it is reasonable to argue that the human perception of the repercussions of breaking a legal rather than social law are generally greater, and that the representations of robots and cyborgs in the key texts studied in this thesis are more likely to have systems based on legally enforceable law as a result of this, nonetheless both forms of 'law' are represented, perhaps most notably in *I, Robot* as the central protagonist is a law enforcement officer, and in *Bicentennial Man* through Andrew's study of family rituals.

The impact of both sides of the nature versus nurture debate result in an immeasurable range of differences across the human species. Such heterogeneity does not negate determinism; event 'y' still has a prior cause. However, such disparity within a species inevitably leads to differences between individuals in the decoding of information - a problem commonly faced by advertising

⁵² Cited by Prigogine, I, in www.youtube.com/watch?v=2NCdpMIYJxQ, *Ilya Prigogine*, accessed Jan 3rd 2010

companies – thus event ‘y’ does not rely solely on event ‘x’ to exist, but may be caused by any number of stimuli depending on the individual concerned. Therefore whilst hard determinism holds that there is no free will, and reality is determined, I consider it important to assert that reality is determined, but no one stimulus will produce the same reaction in all respondents.

This difference can clearly be seen within the opening minutes of *I, Robot* when Will Smith’s central protagonist, Spooner, visits his Grandma. Immediately their contrasting views on robots are made clear to the audience. Though, in order to create enigma, the scene in question does not disclose the reasons for Spooner’s intense negativity towards his Grandma’s hopes of winning a new NS-5 robot in the city lottery, it is ultimately revealed that his experiences of robots differ greatly from those of his relative. Specifically, his misgivings about robot-kind stem from a near-fatal accident which highlighted the lack of compassion exhibited by robots such as the model his Grandma hopes to win – a lack of compassion and therefore response to punishment that Spooner fears.

DETERMINISM VERSUS DUALISM

As a determinist, I have already raised my rejection of dualism in favour of causality. Here I wish to offer my reasons for doing so with specific reference to robots and cyborgs. Dualism, colloquially referred to as the mind/body problem is a conundrum long discussed by prophets and philosophers dating back as far as Zarathustra, Plato and Aristotle. Dualist thinking revolves around the existence of a soul, an incorporeal and eternal element of living beings that is thought to consist of one’s consciousness, personality and wisdom.

It was Rene Descartes’ assertion “Cogito ergo sum” (re-written in *Principles of Philosophy* (1644) from the French version “Je pense donc je suis”, and loosely translated as “I think, therefore I am”) that became popular within western philosophy and formed the basis of, or, for Cartesian philosophers, the solution to many critical ideas and discussions surrounding the nature of human identity and consciousness. Descartes posits the notion of a mind-body dichotomy, a soul clearly distinct from the corporeal which forms the seat of consciousness. This clear definition of the body and soul suggests a relative lack of value attributed to the physical body and negates the possibility of non-human consciousness and personhood through its clear assertion that the soul, containing the consciousness and therefore the essence of a being, is an exclusively human quality.

As discussed previously though, Robert Mitchell and Daniel C. Dennett amongst others make a clear distinction between the person and the human person, and include self-consciousness in their criteria for personhood, thus allowing for the consciousness in non-human species that Descartes' ideas reject. In his essay *Consciousness in Human and Robot Minds*, Dennett offers a series of reasons often advanced for the impossibility of robot consciousness:

- (1) Robots are purely material things, and consciousness requires immaterial mind-stuff. (Old-fashioned dualism).
- (2) Robots are inorganic (by definition), and consciousness can only exist in an organic brain.
- (3) Robots are artefacts, and consciousness abhors an artefact; only something natural, born not manufactured could exhibit genuine consciousness.
- (4) Robots will always just be much too simple to be conscious.⁵³

The first point refers to a distinct separation between mind and body – a state beyond the capacity of a material robot or cyborgs given their constructed and therefore controlled nature. Robots, the first point argues, are comparatively simple objects made of mechanical parts which do not experience anything beyond the material, however much the narratives of the key texts studied herein might suggest that they do. According to the experience of Mitchell and both his and Dennett's criteria though, non-human consciousness is observable in animals such as dolphins, in spite of Descartes.

Sonny's implied level of programming in *I, Robot* seems to afford him qualities that meet the remaining criteria in addition to consciousness. There is an argument for his personhood and that of Marcus Wright, who is not 'purely material'. As a robot, Sonny's consciousness cannot come from 'immaterial minded stuff', but instead from his programming and interaction with his environment in the same manner as Damasio suggests in his Somatic Marker Hypothesis. As such, it is determined by a causal chain of events, and therefore I reject the first point on that basis.

Further, both Descartes' notion and the second and third points raised by Dennett are countered by proponents of epiphenomenalism who instead contend that mental events are caused by physical events and themselves have no physical consequences. Thus as those 'physical events' referred to are chemical processes, the firing of various cerebral neurons through synaptic junctions, they are potentially reducible to fundamental physics. As such, support is given to the possibility of non-human consciousness and personhood. If we accept that humans are conscious beings, whether or not we advocate a dualist philosophical approach, the fact that humans, like robots, are comprised of a series of complex chemical reactions and physical

⁵³ Dennet, D.C., *Consciousness in Human and Robot Minds*, in Schneider, S., (Ed.), 2009, *Science Fiction and Philosophy from Time Travel to Superintelligence*, Ch. 16, p.186-190.

mechanisms suggests that robots too could experience a form of consciousness. Summarised by Daniel C. Dennett,

The best reason for believing that robots might one day become conscious is that we human beings are conscious, and we are a sort of robot ourselves. That is, we are extraordinarily complex self-controlling, self-sustaining physical mechanisms ...operating according to the same well-understood principles that govern all other physical processes in living things.⁵⁴

Of course that 'form' of consciousness I refer to may well be, as Dennett suggests by comparison with a robot's "more elementary artefacts", far less developed than that experienced by a human, but nonetheless the possibility of its existence remains intact and Dennett's fourth and final point is countered. Whilst the potential moral and ethical impact of the acceptance of consciousness in robots and cyborgs is doubtless far-reaching, this thesis remains concerned not with quantification, but with the identification of what elements are necessary for the development of personhood, and of those (physical or otherwise), which determine humanness as opposed to non-human personhood and therefore differentiate us from the representations of artificial identity in the films studied herein. Although not a singular determinant of humanness, consciousness, according to Dennett, is a prerequisite of personhood, and therefore the artificial identity of the non-human protagonists in the key texts. Consequently, we must ask, if consciousness is a determinant of personhood what, in the words of Daniel C. Dennett, is the "special ingredient that is requisite for consciousness?"

REJECTING FREE WILL

Additionally, as a determinist I also do not believe in free will. It is my argument that both humans and non-human animals have the ability to think for themselves, both independently and as part of their own relevant social structures, and therefore operate within what Herbert Simon describes as a "bounded rationality", a "science of the incomplete".⁵⁵ That is to say that we act with knowledge of our own embedding in the world, and thus, to use humans as an example, base our decisions in part on the laws and social constructs (for example gender roles) that we operate within on a daily basis, and the consequences of not doing so. This suggests a degree of restriction of free will, though little more than casual observation of national crime statistics is needed to evidence that the restriction can, of course, be overcome.

⁵⁴ Ibid, p.186

⁵⁵ Prigogine, I, from www.youtube.com/watch?v=2NCdpMIYJxQ, Ilya Prigogine, accessed Jan 3rd 2010

The concept of free will, though, is problematic with regard to how 'free' it really is. Depending on the reason for the breaking of a law, it would be possible to argue that however free the act may appear, if it was coerced in any way it is no more an act of free will than those of individuals who choose to remain within social boundaries. When applying such thinking to the representations of robots and cyborgs upon which this thesis is based, it is clear to see that limitations are placed on their 'choices' by their programming and physical structure – an issue highlighted both by Sonny's denser alloy and Andrew's inability to die.

Ultimately that argument rests on the definition of free will that one subscribes to, be it an act that *could* have been changed if the individual had wanted to do so, or an act that is entirely uncoerced with genuine alternatives, thus placing the responsibility for the outcome squarely upon the individual. For the purposes of this thesis I will adopt the former, given that the representations of cyborgs in the films studied herein are created, 'socialised', trained and judged by the social laws of the humans with whom they interact, thus could be argued to have a limited ability to determine these genuine alternatives due to programming limitations.

Based on assertions like those above, several theorists have advanced that free will is little more than an illusion. Though a number of philosophical and scientific positions on free will have been postulated, in its most basic sense the discussion may be addressed through a study of determinist and indeterminist approaches. The former outlines the predetermination of all occurrences in a universe where each state is possible only in its existing form, as all events are entirely dictated by causal laws.

Prigogine takes an indeterminist stance as he cites Nobel Prize-winning French biologist Jacques Monod's discussion of the scientific conundrum of how to reconcile physics with the evolutionary nature of life, and its inherent unpredictabilities.⁵⁶ Physicalist views stating that nothing can be more than the sum of its physical elements have parallels with reductionist views that, like Dennett's discussion of the creation of a simple brain, discuss the reduction of complex systems to more simplistic versions. Prigogine uses Richard Feynman's metaphor of the world as a game of chess to highlight the incompatibility of the determinist stance with what Prigogine describes as the creativity of the laws of nature.

Prigogine's use of another discipline, art – sculptures in particular – to highlight the temporal qualities that we imbue such work with stands against the eradication of time, a problem experienced acutely by Andrew in *Bicentennial Man*. Frequently, the cyborgs represented in the films around which this thesis is based are shown either considering, or having their 'identities'

⁵⁶ Ibid.

considered by others. Usually this is in relation to human identity, and sometimes by comparison with other cyborgs or robots. Their constructed origins and perceptions of their own bodily and surrounding visual landscapes preoccupy these characters, as Sonny discusses his 'Father' and his difference from other robots, whilst Marcus Wright reflects on his own complex identity and past. In a sense as they are constructed, imagined and realised as opposed to conceived and born, they are arguably sculptures themselves. Prigogine's notion of the temporality of sculpture, then, is applicable to the visual appearance of such representations as Marcus and Sonny in that it demarcates their place in time, and highlights the erosion of the binary opposition discussed herein.

Prigogine continues, contrary to Rousseau's assertion that all nature is a harmony,⁵⁷ chaos is a form of unstable order, not disorder, in which the temporal sequences are very complex. There is, he says, harmony in chaos. Similar possibly to Damasio's contention that emotion and reason, often thought to be unrelated, in fact exist in a facultative symbiosis, and are representative of the experiences of the protagonists in the films studied within this thesis. A lack of equilibrium can produce coherence, structure and complex patterns. Indeed, in his own Dissipative Structure Theory,⁵⁸ focussed on the evolution of complex systems which could include the human body, Prigogine takes an indeterminist step away from Newtonian thinking, arguing that determinism is now outdated and no longer viable.

However, although I accept Prigogine's view that, when studying human beings, a solely determinist view is insufficient, this thesis is based upon the study of representations of robots and cyborgs which, as discussed, are limited by their physical structure and programming in a far greater manner than humans as a species are by their bodies and minds. Evolutionary processes referred to by Prigogine are not relevant to robots and cyborgs, and with regard to the human protagonists studied herein (and cyborgs Spooner and Wright), although I believe all actions to have causes and consequences, I have acknowledged that owing to the heterogeneity of humans and their representations in film, one stimulus will not produce the same reaction in all humans.

But even if we are to endorse the indeterminist, compatibilist views of Prigogine and others, we must also ask, if indeed free will does exist and is a prerequisite for consciousness and thus indirectly for personhood, where, biologically, does it originate in the human body so valued by Damasio, yet relegated to insignificance by Descartes? Are actions of free will rationally based,

⁵⁷ Rousseau, J., Cited in <http://www.youtube.com/watch?v=wbHMc2bO98M> *Ilya Prigogine Chaos Concept*, accessed 12th Feb 2010.

⁵⁸ Cited in *Ilya Prigogine, Chaos and Dissipative Structures*, Research and Development of the U.S. Department of Energy Online, in <http://www.osti.gov/accomplishments/prigogine.html>, accessed 14th Dec 2010.

planned and applied strategies, or are they in part or instead determined by Damasio's somatic markers and therefore linked to a range of bodily organs, including the heart and brain, linked by a complex autonomic nervous system?

The earlier discussion of "bounded rationality" and our awareness of laws, action and consequence suggest the process is a rational one, highlighting the brain as perhaps the most significant organ in determining humanness. Further, this links back to Dennett's arguments proposed regarding the impossibility of robot consciousness based on human inability to create a model as complex as the human brain. But in essence we therefore return to something of a mind/body debate, this time with the focus on the representations of cyborgs in the films mentioned herein.

Each cyborg, be it Sonny, Marcus or Andrew, for example, has a different combination of organic organs and inorganic parts. Andrew in fact 'upgrades' his parts (I use this term both in reference to the privileged position one half of a binary opposition is often afforded, in this case human over robot, and Andrew's ultimate goal) throughout the film to create an increasingly organic version of himself, and the existence or otherwise of his humanness is discussed in terms of which elements have been replaced. The organ synthesis necessary to augment Andrew is carried out by long-time friend Rupert Burns (Oliver Platt) in what is essentially a small warehouse, suggesting the comparative simplicity with which advances that the viewer would consider groundbreaking, could be made.

In *Bicentennial Man* the mise-en-scene constructs a utopian environment that allows for an altogether greater focus on Andrew's struggle to become human. Hover cars fly amongst perfectly uniform steel and glass office complexes – the ultimate in modernity – though not as tall and threatening as the imposing skyscrapers in *I, Robot*. In Andrew's proximity, these contemporary structures are juxtaposed with lakes surrounded by antiquated stone arches, park benches, colourful planted borders and other human augmentations of the natural environment, and it is Andrew's interaction with this environment, be it watering flowers or collecting driftwood, that lead him to want to become a human himself.

However, the intricacy of the mise-en-scene is not reflected in the quality of a narrative which foregrounds scenes designed to generate an emotional response from its human audience, at the expense of addressing the more complex and difficult moral and ethical questions that Andrew's metamorphosis poses. Perhaps, given the human audience's capacities to experience emotion and to identify with the desires Andrew appears to exhibit (I say 'appears' in relation to the discussion of Damasio's Somatic Marker Hypothesis), this narrative construction can be viewed

as successful in the sense that we are perhaps more likely to retain the essence of the film after experiencing a strong emotional response, than if the narrative had paused to reflect on just how a robot was able to open a bank account.

Conversely, we may argue that although legal rights and the nature of man are not necessary considerations when directly applying Dennett's criteria, such a positive representation of humanity as this, which leads Andrew to alter his entire being, problematically ignores the negative behavioural aspects of our species that are foregrounded in *I, Robot* – greed, anger, hate, jealousy – and lacks any real focus on the impact of negative emotions, such as grief. Nonetheless, the film does provide an appropriate platform from which to study the degree to which Andrew does, or does not, meet Dennett's criteria. Additionally, given the ultimate verdict on Andrew's humanity, it is also possible to study what this film suggests is the special ingredient that distinguishes Andrew the android from Andrew the human.

As Richard Martin becomes aware that there is something different about Andrew from all other NDR-114s, he journeys to the headquarters of NorthAm Robotics to seek the opinion of its Head, Dennis Mansky. As Martin describes evidence of 'creativity, curiosity and friendship' in Andrew, presenting some 'carved original works of art' that his robot has made, it becomes clear that Mansky's purpose is to verbalise the fears and nightmares that Nichols discusses in relation to automata. Martin's view of Andrew as 'unique' is seen instead by Mansky as an entirely negative anomaly. 'It has human form, and therefore you read mechanical failure as eccentricity and anthropomorphising', he retorts, in a statement which raises Damasio's Somatic Marker Hypothesis again, in the sense that what Martin, and the viewer sees in Andrew may only be simulation - 'something in the pathways'.

However, in both his carving of these pieces of art, and his reason for doing so (to replace a version he broke), Andrew is both rational and intentional, and perceived as being so by Martin and Lil' Miss. Further, Andrew demonstrates understanding of intent in others, including Mansky's wish to decommission him. As he communicates verbally throughout the film, the majority of which is centred around the 'improvement' of himself, and demonstrates an awareness of time in both his conversation with 'Sir' outside NorthAm, and in his wish to die and no longer be timeless, Andrew satisfies Dennett's criteria for personhood.

His timelessness is, of course, the central barrier to the declaration of his humanity, and in turn his ability to marry Portia. As it is still uncommon for personhood to be thought of outside of the human race, it is unlikely that Dennett's criteria have been applied to automata. Though consciousness is not determined by mortality and thus the ability to die is not a criterion of

personhood, it is an inescapable fact of humanness. Like Sonny's, Andrew's earlier robotic appearance separates him from the humans around him, but in contrast to Spooner's nemesis, Andrew is able to augment his external, as well as internal appearance, aiding the viewer's developing acceptance of him as a person, if not yet a human.

Humanity, then, in *Bicentennial Man* is shown as a jigsaw of parts: "Are you human here?" Andrew is asked by a judge, who motions to his head. The same judge, it is pointed out by Andrew, has a synthetic kidney, yet his humanity (and capacity to judge that in others) is neither compromised nor questioned. The same is true of Detective Spooner in *I, Robot*. Smith's character is partially inorganic as the result of an arm replacement following injury, but continues to act as an enforcer of human law and to be classed as human himself. One would not question a human amputee's level of humanity or consciousness, nor ostracise the wearer of a pacemaker. Therefore it seems clear, certainly from the filmic examples that form the basis of this thesis, that with regard to the notions of humanness, consciousness and free will, certain organs hold greater importance and consequence than others.

EMOTION AND REASON

Although the study of psychopathology and neurophysiology around the 1900s by Sigmund Freud, Charles Sherrington and others, appeared to signal a focus on emotional states unparalleled in earlier years of study that would expand our understanding of their causation, effect and interaction with cerebral processes, the ultimate focus of their work was rather different. Indeed despite the work of Charles Darwin in the late 1800s demonstrating the presence of emotions very similar in essence and causation to those of humans in non-human species (though these were organic species, thus no direct support is lent to the concept of 'artificial identity' in representations of cyborgs from this), cognitive approaches by and large with few exceptions in the twentieth century focussed on the physicalities of brain function rather than emotional research.

As such when internationally acclaimed neuroscientist and neurologist Antonio Damasio wrote *Descartes' Error* in 1994, his assertion that emotion and reason are linked rather than being two quite separate entities was initially met with scepticism. Understandably perhaps as Damasio's proposition was in contradiction to both mainstream scientific thinking and Cartesian dualism. Damasio advanced what he termed the 'Somatic Marker Hypothesis', where each somatic marker demarcates a "gut feeling", an emotion, and "skin conductance responses", the physical feelings

generated by our senses and experiences of our environment, are indices of somatic markers. Damasio argues that emotion is “in the loop of reasoning”, thus where earlier approaches to cognitive activity believed emotion to disrupt the ‘robotic’ reasoning of the mind, or to use our example, Data, instead it supports it, acting as additional information upon which a more considered conclusion can be based.

However Damasio’s thinking, when applied to representations of cyborgs, does highlight some points of differentiation with humans. He allows for the existence of “nonconscious markers” which discriminates between human and cyborgs in that a cyborg is unlikely to be capable of experiencing these given their comparative simplicity to the human brain and body. Further, Damasio states that actions can be based solely on emotion rather than reason – fear responses being an obvious example. Arguably this process is more likely and ingenuous than the reverse as the Somatic Marker Hypothesis states that as our senses are indices of these gut feelings, it is unlikely we would ever be in a situation that we did not experience through one or more of our senses, therefore would not be completely devoid of emotion.

The representations of cyborgs discussed herein, with the exceptions of Detective Spooner and Marcus Wright, begin from a point of programmed reason and aim to move towards a more emotionally capable state. This contradicts Damasio’s assertion that the reasoning ability of the human species “evolved as an extension of the automatic emotional system, with emotion playing diverse roles within the reasoning process.”⁵⁹ Those protagonists in these films that we ultimately judge as human are those who began with a human brain which is the result of the evolutionary process Damasio describes. Those others – Andrew, Sonny – who strive to develop personhood do not have human brains, and can therefore simulate emotion and consciousness and behave in a manner that appears human, but is not classified as such.

If we accept then, with note taken of the issues above, that the representations of cyborgs studied in this thesis can, in principle experience a form of emotion that does not conflict with their programmed, reasoned disposition, albeit perhaps a more limited experience, that emotion must somehow be generated if it is not programmed. How does this generation of emotions occur? This question is partially answered by epiphenomenology and the physical causation of mental effects. These cyborgs are able to sense their environment and their physical interaction with and experience of said environment would then cause a mental event, but without any specific physical reaction, observable if not measurable to indicate emotional experience. As Data does

⁵⁹ Damasio, A., 2006, *Descartes Error*, Preface, p.xxvii-xxviii.

alter his physical reactions seemingly as a result of mental occurrences, there is a clear limitation in this approach which is addressed more closely by Damasio.

Damasio's thinking rejects traditional Cartesian dualism in favour of a more interactionist stance. Though still a form of dualism and similar to epiphenomenalism in that both approaches admit a physical causation, interactionism holds that the relationship between causation and effect is bi-directional; that mind and matter, whilst distinct and independent, exert causal effects on one another. Essentially, whilst epiphenomenologists rely on the aforementioned uni-directional approach to physical and mental causation and effect, it is the bi-directional approach of interactionism that is key here.

Put simply a physical event may result in a mental product and vice versa. Therefore if we apply that theory to the visual context it suggests that our senses are the key to how we experience and interpret the world – cinematic or otherwise. What we see, hear, touch, smell and even taste combines to create a feeling to accompany our physical experience. That feeling in turn informs our next physical experience and so forth. Humans interact with their environments via complex nerve systems that register heat, cold, pain and other sensations. Thus if we return to Dennett's assertion it becomes apparent that robots too are able to detect and interpret their environment, albeit via cruder inorganic systems, hence could be argued to have a form of consciousness. As those systems become increasingly complex, for example in the representations of cyborgs in the contemporary films studied here, so does the depth of their 'consciousness' and resulting 'artificial identity', as we have noted consciousness as one criteria for personhood.

With regard to our senses, arguments I have presented earlier in this thesis note that the human brain is likely to be the key to humanness as it is the irreplaceable element of our organic system which interprets chemical signals from other sensory organs, generating emotions and feeling – Damasio's somatic markers. Obviously therefore without those sensory organs the brain would not have the chemical signals to interpret, and the resulting emotions would not come to be. However I am specifically not suggesting that those same sensory organs are as key to humanness as the brain simply because they are far more easily replaceable with pressure, temperature, light and others sensors, both in humans and robots or cyborgs. Granted those sensors would likely be less complex and effective, but would nonetheless be able to produce enough information for the generation of emotions via the brain.

THE SOMATIC MARKER HYPOTHESIS: DOES CONSCIOUSNESS ALWAYS RESULT IN AN EMOTIONAL RESPONSE?

There is an underlying assumption in this discussion of interactionist thinking, which I would argue is mirrored in filmic representations, that consciousness and senses – an awareness and knowledge of oneself combined with the ability to physically interpret our current environment and any positive or negative stimuli – will result in emotional responses. It is on these grounds that Damasio's Somatic Marker Hypothesis appears to lend support to the possibility of personhood in cyborgs given their limited sensory ability to interact with and interpret their environment. The physical structure and stimuli in their surrounding environment will, according to interactionism and the Somatic Marker Hypothesis, cause a mental event – an emotional response – which itself will result in a physical action and so forth.

However, writing in *Descartes' Error* Damasio presents a case study that is both a challenge to this assumption and further comment on Dennett's assertions noted earlier regarding the impossibility of robot consciousness. Damasio discusses the integration of the autonomic nervous system into both the viscera and musculo-skeletal system of the human – note organic – body. Throughout the body structure run countless innervated blood vessels, each allowing the individual to experience sensations on multiple levels at numerous sites.

A 'gut feeling' can be literally that thanks to the nerve endings that run through the organs located in our abdominal cavity. Though a human stomach cannot be literally tied in a knot by a physical reaction to a stimulus, the concentration of nerve endings in this area can produce a sensation experienced and recognised by humans as such. This allows us not only to interpret our own environment and stimuli, but also to identify with others who have experienced similar sensations. It is these responses to stimuli at visceral and musculo-skeletal levels that form the basis of the Somatic Marker Hypothesis and allow us to modify our physical responses accordingly.

There is further debate linked to the notion of free will and the nature/nurture debate here as to how much influence primary and secondary socialisation as part of a human society has on one's ability to determine what an appropriate physical response to a mental event might be, and it is not a debate that this thesis is focussed on. Damasio notes that 'emotions are required for the proper deployment of social behaviour'⁶⁰ which includes the ramifications of actions of free will, given that the chemical signals from sensory organs which result in emotions are decoded in the brain, the importance of this organ is once again highlighted. I would argue the Somatic Marker Hypothesis should be viewed as a significant, but not singular generator of emotional responses.

⁶⁰ Damasio, A., 2006, *Descartes' Error*, Preface, p.xx.

Other factors such as macro and micro sociological features should be taken into account in order to undertake a study of all factors resulting in the generation of emotions, including the impact on and of the aforementioned peer identifications, on a scale this thesis does not allow for.

Most relevant here to the study of possible personhood is the notion that knowledge and emotion are not inextricably linked. In Damasio's words, '...to know does not necessarily mean to feel.'⁶¹ Along with psycho-physiologist Daniel Tranel, Damasio conducted an experiment to test the Somatic Marker Hypothesis which was based on skin-conductance responses; what we would recognise as the basis of the chat show-popularised 'lie detector test'. When a somatic state is enacted by an individual following a thought – to use our example the mental reaction of a human character in a science fiction film to a cyborg – the autonomic nervous system referred to earlier increases the production of fluids in the sweat glands in human skin.

Though in cases of extreme emotional response (or somatic state) this increase in fluid can be significant, resulting in sweat dripping from the brow or clammy hands, in less extreme somatic states, perhaps those reached through a more indirect experience such as viewing an image rather than directly experiencing the image first hand, the change in fluid can be so slight as to be unnoticeable. However, in the case of a skin conductance test, a harmless electrical current is passed through electrodes attached to the skin of the respondent. The increase in fluid secretion, minimal as it may be, is nonetheless enough to reduce the resistance to the electrical current, and the change in readings are recorded electronically as a wave.

In Damasio's experiment,⁶² the skin conductance responses of individuals with frontal-lobe damage were compared with the responses of individuals with no damage - 'normal' brains. Though all individuals were tested prior to the experiment in order to determine that they could produce skin conductance responses using conditions known to consistently elicit responses in all individuals, there were marked differences in test results between those with frontal lobe damage, and those without. Participants were shown, individually and in the form of a slideshow, a range of images ranging from abstract art and markings to images of homicide and suffering.

Those without damage to their frontal lobes were able to report (and record) clear emotional responses to the images, as well as to provide a verbal description of those emotions and the mental event behind them. However, the individuals with frontal lobe damage did not register skin conductance responses to the negative imagery, despite being able to explain the content of the images and what emotions they should generate with an explanation of their thinking. They were

⁶¹ Ibid, Ch.9, *Testing The Somatic Marker Hypothesis: To Know But Not To Feel*, p.211.

⁶² Ibid, p.204-212.

even able to identify where in the slideshow these images had come in relation to the show as a whole and to other images. Yet their knowledge of the image content did not result in an emotional response.

The implications for this thesis then are two-fold. Firstly, knowledge of a situation or context (for example the visual context of cyborgs referred to herein) does not automatically determine an emotional response in either humans or cyborgs. Secondly, the attribution of functions required for emotional response to specific areas of the human brain may allow for the development of similar inorganic systems that could be incorporated into a cyborg, thus generating a simulated emotional response. Of course, if the cyborg concerned 'feels' that emotion, reacts to it as would a human (or at least records a response identical in nature), it is debatable whether the origin of the emotion should really form a significant part of any argument as to whether the emotion itself is 'real'.

If emotions are 'felt', experienced things, provided there is some evidence of a physical reaction, influenced directly by and proportional to a mental event from the perspective of the individual, that would suggest the emotion exists, or existed. If emotions are that 'special ingredient' mused over by Dennett, and form the root of consciousness and therefore personhood, it is important to challenge this underlying assumption and to note that the possibility of creating a system similar to these specific areas of the brain would carry substantial impact, though it remains only a possibility.

Further, this discussion of emotion has clearly identified the human brain as the site responsible not for the generation of emotions, but as the receptor of information from the sensory points throughout the visceral system that interpret ones surroundings, which ultimately through the brain 'hub', results in action or inaction on the part of the individual experiencing these sensations. Thus as these actions or otherwise are the post-mental event behaviours that we may interpret as emotional, and judge as human, the brain is held as key in this sense to our perceptions of humanity in others. It is my study of these behaviours in the representations of robots and cyborgs in the films noted that form the basis of this thesis; films that also appear to privilege specific organs above others in the formation of the human self, and in the case of said filmic representations, the existence of personhood and artificial identity.

To summarise, although at least a belief in free will if not an actual experience of it forms part of our 'humanness', I believe that this 'special ingredient' that makes us human, rather than which creates non-human personhood cannot be either free will, consciousness or the ability to experience emotion. The existence of true free will, or at least our ability as apparently conscious

and, ideally, morally and ethically sound beings to fully exercise it, is debatable. Likewise consciousness is a condition observable in humans certainly, but also in countless other species, and thus whilst it may be a prerequisite for humanness, does not singularly define it.

Emotions are a regular feature in daily life for humans, drawn from a plethora of feelings in reaction to an even greater range of stimuli. Though the intention of this thesis was never to quantify the number of human emotions by comparison with other species, it seems reasonable to argue that given our social and intellectual development, we probably experience the widest range of emotions of any species. However, even assuming this was a proven fact does not make it the feature I am searching for because the nature of said fact acknowledges the existence of emotions in other species. Regardless of how much narrower the range of emotions experienced by other species may be by comparison to those of humans, some – fear for example – are clearly observable in non-human animals, and thus emotion cannot be the ‘special ingredient’.

The ability to reproduce and to change is not shared by robots and cyborgs. Given their programming and, in the case of the representations of robots and cyborgs in *I, Robot*, Lanning’s three laws, neither is unpredictable behaviour. But of course these are not limited to, and therefore do not define humans. Humour seems to be a uniquely human trait given the understanding of both verbal and non-verbal communication required to affect the correct timing required for a successful joke. Though we may observe playful behaviour in non-human animal species that we interpret as humorous, we are deciding the behaviours according to our own meanings and beliefs in order to find it amusing, rather than perceiving it in the manner that other members of the receiving group would. As such humour does seem uniquely human. But the timing and knowledge required to make a joke work are acquired through our social interaction with our own species and environment at a sensory level and interpreted and stored by the brain.

Instead I believe the essence of humanness is in the irreplaceable nature of the human brain. Earlier discussions in this thesis outlined both our lack of technological knowledge and financial ability to accurately replicate this organ – an organ which acts as the central hub through which all sensory information is passed and interpreted, in turn generating aspects of the mind including emotion, will and self-awareness. Replace a human’s limbs, kidneys, liver or lungs and we still consider them human. Even the heart can be mechanically controlled via a pacemaker or replaced with that of another human or even an animal, and the question would be how best to enjoy the resulting new lease of life, rather than to debate its nature. But if a human’s brain were

to be replaced with a mechanical or other non-human alternative, would the recipient still be considered human? I believe not.

Damasio discusses the notion that ‘...brain systems that are jointly engaged in emotion and decision-making are generally involved in the management of social cognition and behaviour.’⁶³ Thus the brain is the root of the actions and feelings which, Damasio continues, ‘form the base of what humans have described for millennia as the soul or spirit.’⁶⁴ Far from Descartes’ assertion that body and soul are distinct, they, like emotion and reason are rooted in each other.

Indeed in *I, Robot* Dr. Lanning quotes Gilbert Ryle’s disparaging description of Descartes’ mind-body dualism, the ‘ghost in the machine.’⁶⁵ In his 1967 book of the same title, Arthur Koestler discusses the evolution of the human brain and the manner in which it has built upon the same more primitive structures that Damasio outlined, not erasing them, but merely suppressing them until such times, for example highly emotional times, when the earlier structures come to the fore in place of those more developed, rationally-based structures. The impulses that accompany the use of these earlier structures – usually negative, destructive emotions and desires – are the ‘ghost in the machine’, and both describe the random differences between characters in texts such as *I, Robot*, as well as accounting for the unpredictability of human behaviour that Rupert calls ‘doing something very silly’ in *Bicentennial Man*. Again, the significance of the brain is raised.

The organic debate as a definition of humanity can even be seen in contemporary pop-culture references, for example the advertising campaign for the Alfa Romeo Giulietta (2010), created by Leo Burnett: “Without heart we are only machines.” Heart in this sense refers to the human emotional connotations of the heart rather than to the organ itself, inferring the process of meaning transfer through the acquisition of the vehicle. After all, the Giulietta does not physically have a heart, therefore as those emotions result from physical interaction with the car, interpreted and generated (as it is the origin point of instructions for the body to react physically to mental events which in turn lead to further emotions) by the brain, the advert does little more than highlight the importance of shared meanings amongst humans – in this case connotations of the heart - in our daily understanding and interaction, which is not shared by robots. Of course, ‘without sensory organs embedded in a massively innervated musculo-skeletal system and a

⁶³ Damasio, A., 2006, *Descartes’ Error*, Preface, p.xix

⁶⁴ Ibid, Introduction, p.xxvi

⁶⁵ Ryle, G., 1949, *The Concept of Mind* in <http://great-philosophers.suite101.com/article.cfm/gilbert-ryle-philosophy>, accessed Sept 12th 2010.

brain to interpret the biochemical messages resulting in positive emotions, we are only machines' doesn't make for such an appealing slogan.

The shared meanings I refer to here are the result of nurture, rather than nature, and of our life-long immersion in the ideology of our own society and culture. As a result of this process, humans from the same societies and cultures often share collective (albeit that term increasingly carries connotations of robots in popular culture) memories and identities. This is not to say that we are, like the standard NS-5s in *I, Robot*, identical. But they are a product of similar developmental environments, and therefore share some similarities in the way they will interpret certain situations, such as the implied meaning in the aforementioned advertising campaign. The fact that, certainly in the Anglo-American societies that are responsible for the texts studied herein, such hegemonic discourses are determined by the dominant, capitalist minority here is irrelevant. Simply the fact that the vast majority of humans within that culture adopt said beliefs as their own is enough to explain the second mechanism of Damasio's somatic marker process – a process based on vicarious as opposed to real body states.

Damasio discusses our social 'tuning' – essentially the impact of primary and secondary socialisation throughout our early life, and argues that the majority of our decision making 'was shaped by somatic states related to punishment and reward'⁶⁶ in a similar manner to operant conditioning. During the experience of a 'real' body state, specific areas of the brain (the amygdala and prefrontal cortices) engage the human body to affect a particular state – a feeling for example. The resulting chemical messages which demarcate that feeling are then 'signalled to the somatosensory cortex, attended, and made conscious.'⁶⁷

However, the second mechanism of the somatic marker process relies on our maturation throughout our childhood, and the repetition of specific positively and negatively reinforced situations that accompanied it. As a result of those events, some of our decision making strategies rely far less on somatic states (i.e. the importance of the body is far less than the brain here, though not completely negated) and instead may be based on 'symbols of somatic states'.

This process of course is one that robots are not involved with as they are not socialised, nor can they be rewarded or punished; the latter being one of the factors noted by Johann Hari upon which human fear of robotic take-over rests. So whilst not exclusively human, the process and its outcomes are a point of differentiation between human and robot. However, Damasio's 'as if'

⁶⁶ Damasio, A., 2006, *Descartes' Error*, Ch.8., *The Somatic Marker Hypothesis - Somatic Markers: Theater in the Body or Theater in the Brain?*, p184.

⁶⁷ *Ibid*, p.184.

symbols appear more robotic than the experience of a real body state as they rely less on feelings. Nonetheless the importance of specific areas of the brain in both sides of this process serve to highlight the overall importance of the organ itself, as well as bring us back to Dennett's ideas of humans as a 'sort of robot', and the possibility that, whilst we are currently incapable of replicating an entire, functioning human brain, perhaps specific elements of it may be closer to our grasp.

Damasio emphasizes the importance of chemical signals in the brain being attended. Without such attention the resulting feelings will never become part of one's consciousness, but despite this whether resulting from a real or vicarious states, those signals can still be part of mechanisms, of which we remain unaware, that govern our approach to the world. Therefore Damasio goes on to discuss the impact of this on organisms with limited brain capacities that do not allow for consciousness, specifically bees in his example, though there is relevance to robots as well.

The covert mechanisms he says are 'a means to build predictions of outcome and bias the organism's action devices for behaving in a particular way, which may appear to the external observer as a choice.'⁶⁸ With reference to robots like Andrew in *Bicentennial Man* then, the behaviours 'he' exhibits which appear to the audience to demarcate the transition from robot to human may in fact be nothing more than predictions of likely outcome based on generations spent with the same family, throughout which time his programming recorded specific events and caused Andrew to act according to calculations based on these.

Our development and implementation of the types of technology discussed by Pilcher, Hari and others now means that both the technology we encounter in everyday life and the representations of it we see in film are more expected by an audience and less extraordinary to them, allowing the kind of man and machine fusion we see in Detective Spooner and others to be readily accepted and frequently presented. 'Our minds would not be the way they are if it were not for the interplay of body and brain during evolution', states Damasio, highlighting the ongoing, developmental nature of our acquisition of knowledge and understanding, of which the evolutionary process of encephalisation has become symbolic.

Therefore, '...the human brain and the rest of the body constitute an indissociable organism'⁶⁹ which 'interacts with the environment as an ensemble'⁷⁰ Damasio continues. As a result, it would

⁶⁸ Ibid, p.185.

⁶⁹ Ibid, Introduction, p.xxvi.

⁷⁰ Ibid, Introduction, p.xxvii.

not be accurate to state that the brain alone is the root of the erosion of the binary opposition between human and robot, as the brain's current form and function have been determined by its interaction with the rest of the human visceral system, which in turn has continually interacted with our changing visual landscapes via our sensory registers to produce information.

The development of the human brain in this sense is symptomatic of the erosion of the binary opposition between human and robot, and as the ultimate point of 'storage' for the knowledge that has allowed us to develop technology, including the technology that has created and is represented in the filmic worlds being studied in this thesis, it is frequently highlighted in such texts as the defining factor in what makes us human. Whilst the resulting similarity between 'us' and 'them' can be unnerving and played upon by a film director, the brain also acts as the differentiating characteristic that permanently separates viewer from subject – an organic cushion of sorts that provides a guarantee of irreplaceability as we explore these representations of dystopic possibilities.

CONCLUSION

Clearly, then, and with acknowledgement of Damasio's Somatic Marker Hypothesis in the sense that simulation may account for some or all of the emotions we recognise in the robotic protagonists of the key films upon which this thesis is based, Sonny, Andrew and Marcus can all be considered persons. But what is the difference between personhood and humanness, called 'that special ingredient' by Dennett? In Sonny's case, there is no attempt on the part of the filmmaker to suggest that he be considered human. Instead, we are asked to reflect upon what makes us human, by comparison to him. Marcus and Andrew, though, are different. The former begins life as a human, and the latter ends his as one. Therefore, we must attempt to differentiate between applications of the term 'consciousness', in order to determine if these two possess any of the characteristics that separate human from non-human.

In his article on human consciousness, which is based on Antonio Damasio's latest book,⁷¹ Steven Rose makes the distinction between the quantitative and qualitative differences between the consciousness of humans, and that of other animal species. Consciousness is, he states, an 'evolved property with a function of some benefit to its possessor.'⁷² As such, Rose is differentiating between the various definitions of consciousness. In its most basic sense, consciousness is discussed as simply a form of awareness; a state where the mental faculties of

⁷¹ Damasio, A., 2010, *Self Comes to Mind: Constructing the Conscious Brain*, Heinemann.

⁷² Rose, S., 2011, *The Mind's I*, in *Saturday Guardian*, 12th February 2011.

an individual are active, but operating at a level where intentional, motivated behaviour is indiscernible.

In terms of Metzinger's notions of globality and presentationality, the individual would not be, even if only temporarily, aware of the integrated world model of which they are a part, or of the 'continuous flow of physical time'. This interpretation may be crudely applied to an individual who has an awareness of themselves and their surroundings, but no awareness of temporality, for example humans with mental disabilities, or who are undergoing surgery through which they remain conscious, though sedated.

Our understanding of consciousness in humans other than those noted above, though, is, in Rose's words, something that 'gives us humans the capacity to learn from the past, to anticipate and plan for the future, to establish and maintain social relations, to imagine and create societies, technologies, art and literature.' He continues to discuss how, via an analysis of the evolution of the homeostatic system of multicellular organisms, Damasio highlights the capacities of brain neurons with myriad connections (specifically those found in humans) to 'form representations of both the external world and the body state of the organism that owns them.' Rose notes that 'such brains enable their possessors to learn and remember, to recognise the present in the context of the immediate past and the imminent future.' Clear links to Dennett's criteria for personhood can be drawn here.

The impact of time (or in some cases, lack of it) on the protagonists, raised in earlier discussions in relation to Locke and Metzinger, is focused on in each text – the changes in Marcus Wright's humanity from organic, to cyborg and through his time with The Resistance; Spooner's prosthetic arm which in itself is not only inorganic and thus not subject to the decay of the rest of his body, but which also forms a link to a specific, traumatic time in his life that determines his current physical and mental state; our study of Andrew across the generations of the family he serves, and his ultimate recognition of time – his immortality – as a factor restricting his ability to be perceived as human.

According to Patrick Fuery, 'Skin is the defining surface of the body and, consequently, the subjectivities; it designates age, race, life style, even class.'⁷³ Andrew's though does not. Rather it stands as a clear marker of his status as a robot, a technological and, given the improved resemblance to humans on the newer NS-5 model, artistic creation, linking to Prigogine's thoughts on art and the temporality of sculpture. Andrew's skin does not age, nor does it

⁷³ Fuery, P., 2000, *New Developments in Film Theory*, p.77

represent class or lifestyle. It demarcates the time in which he was constructed, and continues to highlight his differences from the humans around him through his agelessness.

In turn as we move towards a discussion founded upon mortality, our ability to die, we therefore must address the issue of the organic versus the inorganic as we began to earlier in the thesis and ask, in light of questioning how a robot might identify a human, what the impact of a partially inorganic human - an amputee with a prosthetic limb like Spooner, or the wearer of a pacemaker, for example - might have on its ability to do so?

However, it is further clarification from Damasio on the subject of the brain – argued throughout this thesis to be the differentiating factor between humanness and personhood – that offers perhaps the best definition of Dennett's 'special ingredient', even if it raises further issues with regard to the measurement of the same:

In animals with big brains, emotions - mere bodily responses - become translated into feelings, and with feelings, a mind - a subtle flowing combination of actual images and recalled images in ever-changing proportions emerges from the brain. Many large-brained creatures thus have minds, however alien they may be to our own. But consciousness only emerges when [...] self comes to mind.⁷⁴

This distinction between brain and mind is made clearer by Damasio's identification of an 'autobiographical memory', made possible, he argues, by sensory experiences from the visceral system, interpreted by the brain, interacting with 'the encoded experiences of the past that self provides.' The result he describes as 'the narrative of our lives that we humans all possess and which is the basis for consciousness.' Of course, the methodology we might employ to attempt to measure a dog's understanding of the narrative of its life would undoubtedly be problematic, and any results flawed. The relative size of a dog's brain compared to that of a human though, and its comparative development in terms of neural structure would suggest this 'narrative' to be far less developed, and therefore the level of consciousness less.

But 'brains are not conscious; people are,' notes Rose. Citing anthropologist Tim Ingold, he continues, 'our brains enable our consciousness, just as our legs enable our walking. But to attribute the property of a whole to that of a part is to commit [...] the mereological fallacy.' Ultimately, both Dennett's criteria for personhood and Damasio's distinction between the brain and the mind do highlight the brain as the organ which most determines humanness, thanks to its size and development. However, although we may hail the brain as the key to humanness thanks to the abilities it affords us, and we can argue that without it, none of these abilities would be

⁷⁴ Ibid.

possible, we should also note that the brain alone cannot generate the range of emotions, feelings and level of consciousness that Dennett, Damasio and others label as human. Mechanical or biological, the rest of the visceral system has a significant part to play in support of that 'special ingredient'.

Both Andrew and Marcus do display a clear awareness of the narrative of their own lives. Andrew's story is literally about his life, from one generation of his family to the next, and the changes that he experiences – physical and mental – throughout his life. Each change he makes is deliberately compared to his status in his immediate past, and with a view to shaping his future to create a new self. Notably, in a thesis which highlights the brain as the ultimate key to humanness, there is no direct reference to a human brain ever being created for Andrew, though a three-dimensional image in Rupert's workshop does appear to show one. Instead, it is referred to as a central nervous system, further highlighting the interaction of the brain with other sensory organs in the generation of emotion and feeling.

Though Marcus's life runs in the opposite direction to Andrew's, he too is seen, through a series of flashbacks, memories and realisations – specifically his appalled reaction to the revelation that his body has been altered after his death and is no longer human – are evidence of his knowledge of his past, just as his final sacrifice indicates his awareness of the future. This final act does point to the heart as being the saviour of humanity of course, rather than the brain, but it was Marcus's understanding of the importance of John Connor to the human race which led to his sacrifice – understanding which occurs in the brain.

Again, it is not clear whether or not Marcus's brain has been replaced with a mechanical version, and some of his human memories simply implanted. His flashbacks and ability to act freely would suggest not, though his 'programming', designed to ensure he gets to Connor imply the opposite. This ambiguity does allow the audience the freedom to believe that both Andrew and Marcus have organic brains, particularly as such acute understanding of temporality is not shown by Sonny. Thus Andrew and Marcus may be considered human, or at least more human, than him.

The robots and cyborgs studied in this thesis are mere filmic representations of the actual versions our 'big brains' are capable of creating in reality, and the films are our way of exploring the possibilities, limitations and dangers of creating such beings. Despite our technological and scientific abilities allowing us to create the stuff of dreams, Nichols also highlighted the nightmares that such skill might create – nightmares explored in detail in dystopian film scenarios like those seen in *Terminator Salvation*. Given our fear of these nightmares, so strong that we have produced film after film, narratives spanning decades which represent our anxieties about

the dangers our creations present, will we continue develop our technology to the point where science fiction becomes reality? In the words Rupert Burns used to describe the human condition to Andrew, that would be 'something very silly' indeed.

BIBLIOGRAPHY

- Alexander, B., 2000, *Don't Die, Stay Pretty: Introducing the Ultrahuman Makeover*, in <http://www.wired.com/wired/archive/8.01/forever.html> , accessed 6th Dec 2009.
- Asadi, A. (Ed.), 2010, *SciFiNow*, Issue 37, Image Publishing.
- Asimov, I. 1978, *The Machine and the Robot*, cited in Warrick, P.S., Greenberg, M.H. and Olander, J.D. (Eds.), 1978, *Science Fiction: Contemporary Mythology*, Harper and Row.
- Biderman, S., *Recalling the Self: Personal Identity in Total Recall*, in Sanders, S.M., (Ed.), 2008, *The Philosophy of Science Fiction Film*, p. 39-54, Kentucky Press.
- Bostrom, N., 2005, *In Defense of Human Dignity*, p.1, in <http://www.nickbostrom.com/ethics/dignity.html> , accessed July 2011.
- Bould, M., Butler, A.M., Roberts, A. and Vint, S., 2009, *The Routledge Companion to Science Fiction*, Routledge.
- Bould, M. and Vint, S., 2011, *The Routledge Concise History of Science Fiction*, Routledge.
- Bowers, C., *How Close is Terminator-Like World?*, from <http://www.cbsnews.com/stories/2009/05/25/earlyshow/main5037912.shtml> , accessed 30th June 2009.
- Brandt, R., (Ed.), 1962, *Social Justice*, Prentice Hall.
- Cave, P., 2007, *Can a Robot Be Human?*, Oneworld.
- Clarke, D.B., (Ed.), 1997, *The Cinematic City*, Routledge.
- Cook, J. And Wright, P., 2005, *British Science Fiction Television: A Hitchhiker's Guide*, I.B. Tauris.
- Curran Trussell, R., *I, Robot: You Gotta Have Heart* in *The Journal of Religion and Film*, October 2005, Vol 9., <http://www.unomaha.edu/jrf/Vol9No2/TrussellRobot.htm> , accessed 3rd August 2010.
- Damasio, A., 2006, *Descartes' Error*, Vintage.
- Damasio, A., 2004, *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*, Vintage.
- Damasio, A., 2000, *The Feeling of What Happens: body, emotion and the making of consciousness*, Vintage.
- Dennet, D.C., *Consciousness in Human and Robot Minds*, in *Science Fiction and Philosophy From Time Travel to Superintelligence*, 2009, Schneider, S., (Ed.), Ch. 16, Wiley-Blackwell.
- Fuery, P., 2000, *New Developments in Film Theory*, Macmillan.
- Gibbs, J., 2002, *Mise-en-Scene Film Style and Interpretation*, Wallflower.

- Hari, J., 2010, *The Age of the Killer Robot Is No Longer a Sci-fi Fantasy*, in <http://www.independent.co.uk/opinion/commentators/johann-hari/johann-hari-the-age-of-the-killer-robot-is-no-longer-a-scifi-fantasy-1875220.html>, accessed 5th Jan 2010.
- Hayward, J., Cardinal, D., and Jones, G., *Are There Any Non-Human Persons?* in http://www.alevelphilosophy.com/index.php?option=com_content&view=article&catid=54%3Aessays&id=48%3Aare-there-any-non-human-persons&Itemid=66, accessed 8th Jan 2011.
- Hill, J. and Church Gibson, P., (Eds.), 1998, *The Oxford Guide to Film Studies*, Oxford University Press Inc.
- Holland, S., 1995, *Descartes Goes to Hollywood: Mind Body and Gender in Contemporary Cyborg Cinema*, in *Body & Society Online Journal*, Vol. 1, No. 3-4, p.157, http://bod.sagepub.com/cgi/pdf_extract/1/3-4/157, accessed 4th July 2009.
- Holt, J., *Terminator – Fear and the Paradox of Fiction*, in Sanders, S.M., (Ed.), 2008, *The Philosophy of Science Fiction Film*, p. 135-152, Kentucky Press.
- Iredale, M., 2008, *Sci-Phi: Claude Levi-Strauss lived for a century and became an icon of one*, in *The Philosopher's Magazine*, Issue 49, 2nd Quarter 2010, p.33-35.
- Iredale, M., 2008, *Sci-Phi: Putting Descartes before the horse*, in *The Philosopher's Magazine*, Issue 42, 3rd Quarter 2008, p.40-41.
- Knight, D. and McKnight, G., *What is it to be Human? Blade Runner and Dark City*, in Sanders, S.M., (Ed.), 2008, *The Philosophy of Science Fiction Film*, Ch.1, p.21-37, University Press of Kentucky.
- Kaveney, R., 2005, *From Alien to the Matrix: Reading Science Fiction Film*, I.B. Tauris.
- Lee, K., 1999, *The Natural and the Artefactual: The Implications of Deep Science and Deep Technology for Environmental Philosophy*, Ch. 2, p.50, in http://books.google.co.uk/books?id=ZeH62s45c3oC&pg=PA49&lpg=PA49&dq=The+Natural+and+the+Artefactual&source=bl&ots=KQ-PMNYvFc&sig=d0qZ6cSSG7T7me8F3LMJ0vJe4nY&hl=en&ei=1Z-8S6WqEY3QmqPT0tnCCA&sa=X&oi=book_result&ct=result&resnum=1&ved=0CAgQ6AEwAA#v=onepage&q&f=false, accessed 7th April 2010.
- Locke, J., 1689, *An Essay Concerning Human Understanding*, Ch.XXVII, *Of Identity and Diversity*, in http://oregonstate.edu/instruct/phi302/texts/locke/locke1/Book2c.html#Chapter_XXVII, accessed 24th Feb 2011.
- Lovgren, S., 2005, *Chimps, Humans 96 Percent the Same, Gene Study Finds*, *National Geographic News*, in http://news.nationalgeographic.com/news/2005/08/0831_050831_chimp_genes.html, accessed 5th August 2010.
- McKibben, B., 2004, *Enough: Staying Human in an Engineered Age*, H. Holt.
- McMahon, J.L., *The Existential Frankenstein*, in Sanders, S.M., (Ed.), 2008, *The Philosophy of Science Fiction Film*, p. 73-90, Kentucky Press.
- Metzinger, T., 2003, *Being No One: The Self-Model Theory of Subjectivity*, The MIT Press.
- Midgley, M., 1998, *Animals and Why They Matter*, University of Georgia Press.

- Mitchell, R.W., 1993, *Humans, Nonhumans and Personhood*, in <http://www.animal-rights-library.com/texts-m/mitchell01.htm>, accessed 19th November 2010.
- Newman, S.A., 2003, *Averting the Clone Age: Prospects and Perils of Human Developmental Manipulation*, in http://genetics.live.radicaldesigns.org/downloads/200303_jchlp_newman.pdf, accessed 7th April 2010.
- Nichols, B., 2002, *Introduction to Documentary*, Indiana University Press.
- O'Brien, D. and Newman, K., 2000, *SF/UK: How British Science Fiction Changed the World*, Reynolds and Hearn Ltd.
- Parthan, B., *Technology – THE CYBORGS ARE COMING*, from <http://www.lifepositive.com/Mind/evolution/technology/cyborg.asp>, accessed 4th July 2009.
- Perry, J. (Ed.), 1975, *Personal Identity*, University of California.
- Pilcher, P., 2009, *Half Man, Half Machine: The Cyborgs Are Coming*, in <http://www.independent.co.uk/life-style/gadgets-and-tech/features/half-man-half-machine-the-cyborgs-are-coming-1822739.html>, accessed 5th Jan 2010.
- Pisters, P., 2010, *The Brain is the Screen: The Neuro-Image in Contemporary Cinema*, in <http://www.sbs.ox.ac.uk/centres/insis/neurosociety/Documents/pisters-neurosociety.pdf>, accessed 31st December 2010.
- Pisters, P., 2006, *The Spiritual Dimension of the Brain as Screen Zigzagging from Cosmos to Earth (and Back)*, in <http://home.medewerker.uva.nl/p.p.r.w.pisters/bestanden/Screen%20Consciousness%20article%20as%20published%20met%20referentie.pdf>, accessed 31st December 2010.
- Pratchett, T., 2010 (Ed.), *SFX*, Issue 196, Future Publishing.
- Prigogine, I., 1997, *The End of Certainty*, Free Press.
- Prigogine, I, in www.youtube.com/watch?v=2NCdpMIYJxQ, Ilya Prigogine, accessed Jan 3rd 2010.
- Pyle, 1993, in <http://www.megaessays.com/viewpaper/39686.html>, accessed 12th July 2009.
- Rachels, J., 1999, *Created from Animals: The Moral Implications of Darwinism*, Oxford University Press.
- Relph, S. and Foster, P., 2007, *Twin Girl With Eight Limbs to Have Surgery*, in The Telegraph online, in <http://www.telegraph.co.uk/news/worldnews/1568406/Twin-girl-with-eight-limbs-to-have-surgery.html>, accessed 5th January 2011.
- Richardson, K., in <http://www.culturewars.org.uk/2004-02/irobot.htm> accessed 7/4/10.
- Roberts, A., 2005, *Science Fiction: the New Critical Idiom*, 2nd Ed., Routledge.
- Rose, C. S., 2007, *Biology in the Movies: Using the Double-Edged Sword of Popular Culture to Enhance Public Understanding of Science*, *Journal of Evolutionary Biology*, Issue 34, p.49, in <http://resources.metapress.com/pdf-preview.axd?code=3016t2154360t6m5&size=largest>, accessed 31st December 2010

- Rose, S., 2011, *The Mind's I*, in *Saturday Guardian*, 12th February 2011.
- Rousseau, J., Cited in <http://www.youtube.com/watch?v=wbHMc2bO98M>, *Ilya Prigogine Caos Concept*, accessed 12th Feb 2010.
- Rugg, G. and Gerrard, S., 2008, *Old Problems, New Thinking*, in *The Philosopher's Magazine*, Issue 42, 3rd Quarter 2008, p.52-57.
- Ryle, G., 1949, *The Concept of Mind* in <http://great-philosophers.suite101.com/article.cfm/gilbert-ryle-philosophy>, accessed Sept 12th 2010.
- Shanahan, T., *The Evolutionary Indeterminism Thesis*, in *BioScience* (2003), Vol 53., No. 2, p163 in <http://www.jstor.org/pss/1314330>, accessed 29th October 2010.
- Singer, P., *All Animals Are Equal*, in Regan, T., and Singer, P., (Eds.), 1989, *Animal Rights and Human Obligations*, New Jersey, p148-162, in <http://www.animal-rights-library.com/texts-m/singer02.htm>, accessed 24th February 2011.
- Skinner, M., in http://hrsbstaff.ednet.ns.ca/mskinner/Sociology/Socialization/feral_children.htm, accessed January 5th 2011.
- Taylor, C., 1989, *Sources of the Self: The Making of the Modern Identity*, Ch.1, *Identity and the Good: The Self in Moral Space*, Press Syndicate of the University of Cambridge.
- Telotte, J.P., 2008, *The Essential Science Fiction Television Reader*, The University Press of Kentucky.
- Templeton, D., 2004, *Bot and Souled*, in <http://www.metroactive.com/papers/sonoma/07.28.04/talk-pix-0431.html>, accessed 4th November 2010.
- Vaknin, S., *On Being Human*, in <http://samvak.tripod.com/human.html>, accessed 8th June 2010.
- Vaknin, S., *The Fourth Law (of Robotics) Part I: The Simple Problems*, in <http://samvak.tripod.com/robot.html>, accessed 8th June 2010.
- White, T.I., 2007, *In Defense of Dolphins: The New Moral Frontier*, Blackwell.
- White, T.I., 2010, *Dolphin People*, in *The Philosopher's Magazine*, Issue 49, 2nd Quarter 2010.
- Wickens, A., 2000, *Foundations of Biopsychology*, Ch.5, *Emotional States*, Prentice Hall.
- www.scifinow.co.uk, accessed 8th June 2010.
- www.sfx.co.uk, accessed 8th June 2010.

FILMOGRAPHY

Aliens, 1986, Dir. James Cameron.

Amistad, 1997, Dir. Steven Spielberg.

Bicentennial Man, 1999, Dir. Chris Columbus.

Blade Runner. 1982, Dir. Ridley Scott.

Dark City. 1998, Dir. Alex Proyas.

Eternal Sunshine of the Spotless Mind. 2004, Dir. Michel Gondry.

Fight Club. 1999, Dir. David Fincher.

Frankenstein. 1994, Dir. Kenneth Branagh.

Forbidden Planet, 1956, Dir. Fred M. Wilcox.

He-Man and the Masters of the Universe, 1983-85, Filmation, USA.

I, Robot, 2004, Dir. Alex Proyas.

Masters of the Universe vs. The Snake Men, 2002, Episode 33: *Of Machines and Men*, Halperin, M., original air date 29/11/03, Mike Young Productions.

Metropolis, 1927, Dir. Fritz Lang

Robocop, 1987, Dir. Paul Verhoeven.

Schindler's List, 1993, Dir. Steven Spielberg.

Star Wars, 1977, Dir. George Lucas.

Strange Days, 1995, Dir. Kathryn Bigelow.

Surrogates, 2009, Dir Johnathan Mostow.

Terminator Salvation, 2009, Dir. McG.

The Terminator, 1984, Dir. James Cameron.

Total Recall. 1990, Dir. Paul Verhoeven

Visionaries Knights of the Magical Light, 1987, Episode 4: *The Price of Freedom*, Sunbow Productions.

Wall-E, 2008, Dir. Andrew Stanton.