ance of the sacred, spiritual
prehension and of the cos-
for good and ill; and of the
ative to provoke our cultural
ons of identity, purpose and

37 The Ghost in the Machine: Humanity and the Problem of Self-Aware Information

Brett Lunceford

Theories of posthumanism place considerable faith in the power of information-processing. Some foresee a potential point of self-awareness in computers as processing ability continues to increase exponentially, while others hope for a future in which their minds can be uploaded to a computer, thereby gaining a form of non-corporeal immortality. Such notions raise questions of whether humans can be reduced to their own information-processing: Are we thinking machines? Are we the sum of our memories? Many science fiction (SF) films have grappled with similar questions; this chapter considers two specific ideas through the lens of these films. First, I will consider the roles that memory and emotion play in our conception of humanity. Second, I will explore the question of what it means to think by examining the trope of sentient networks in film.

Every cell in our bodies contains information that dictates what we are, and small changes in this information can yield drastic changes. The four nucleobases of our DNA could be rearranged to form other entities; our DNA differs from that of chimpanzees by only 1.24 per cent (Ebersberger et al. 2002). In some ways, we *are* the information contained in these intertwined strands, but in other ways we are not. Even if I were to create a clone of myself, I would still not be able to endow that individual with my idiosyncratic outlook on life, sense of humour, memories or feelings, because those are influenced by the experiences that I have had over the course of an entire lifetime, and the chronology and interplay of these events. Although these things are information that I can communicate to others, I may be limited by vocabulary with which to express these emotions and experiences, or even by my own memory. We are limited by our humanity. As Hauskeller (2012) notes, the gulf between copying information and copying an exact duplicate of a *self* is enormous. But what if we could overcome these limitations through the use of technology? What if we were able to figure out how to truly operationalize, encode and decode emotions and memories? What would that mean for our humanity? Do we need humanity to feel human emotions?

This chapter explores two major themes. First, I will consider how memory plays a key role in our conception of what it means to be human. I will argue that the focus on memory and emotion reinforces a Cartesian split between the mind and body – a belief often found in the literature surrounding posthumanism.

Second, I will examine the question of sentience and its role in our understanding of humanity. The very notion of sentient technologies challenges the perceived human monopoly on self-awareness, and film portrayals of those systems that achieve sentience often do more to describe the present human condition than to forecast some potential technological future.

Although popular film is firmly rooted in the realm of fiction, such narratives have significant consequences. As Vint (2007, 172) suggests,

> *Struggles over posthuman identity are thus more than struggles over technology and the ethics of various technological ways of modifying what it means to be human. Rather, they are also and more importantly political. The category of the human has historically been used in exclusive and oppressive ways, and the category of the posthuman entails similar risks.*

As such, one cannot simply dismiss these films because they are fiction. The stories we tell often say more about our own values, thoughts and anxieties than about the protagonists.

Theories of posthumanism, or humanity isn't what it used to be

The terms posthumanism, transhumanism and humanity+ are often used interchangeably, and, although there are differences (see Krueger 2005), for the purposes of this chapter I will focus on what these philosophies have in common and use the umbrella term 'posthumanism' to encompass them all. However, in order to explain posthumanism, we must first begin with the question 'which posthumanism?' There seem to be two different camps – academics and popular culture. These two camps are not mutually exclusive, but it is important to recognize the frameworks in which they operate. For academics, posthumanism is a natural outgrowth of postmodernism, in which the assumed Enlightenment subject is called into question. According to Hayles (1999, 3), 'the posthuman subject is an amalgam, a collection of heterogeneous components, a material-informational entity whose boundaries undergo continuous construction and reconstruction'. Put another way, Pepperell (2003, 171) explains that posthumanism 'is about the end of "humanism", that long-held belief in the infallibility of human power and the arrogant belief in our superiority and uniqueness'. Pop-culture posthumanism, on the other hand, 'envisions the challenges to the human as largely corporeal ones resulting from our supposedly intractable situatedness in the so-called natural world' (Seaman 2007, 248). As one may expect, pop-culture posthumanists have attracted much more attention, in part because of the practical aspects of their predictions and their perceived outlandishness.

Adherents to these different strands of posthumanism wish to transcend and reconfigure humanity for different ends, but all see technology as the catalyst that will allow us to overcome the frailties and weaknesses of humanity. Graham (2002, 69) explains, 'The philosophies and practices of transhumanism exhibit a will for transcendence of the flesh as an innate and universal trait, a drive to overcome physical and material reality and strive towards omnipotence, omniscience, and immortality.' Still, some seem to see this desire to extend ourselves through technology as an intrinsically human attribute rather than a posthuman one (McLuhan 1994). Clark (2003, 142) argues that 'such extensions should not be thought of as rendering us in any way posthuman; not because they are not deeply transformative but because we humans are naturally designed to be the subjects of just such repeated transformations.'

Another commonality these philosophies share is a deterministic view of technology. Nicholas Negroponte (1995, 231), for example, argues that we are all becoming digital: 'It is almost genetic in its nature, in that each generation will become more digital than the preceding one.' In many of these writings, agency becomes difficult to locate as machines are granted a kind of will. Haraway (1991, 177) suggests that 'it is not clear who makes and who is made in the relation between human and machine'. Perhaps one reason why the lines of agency are becoming blurred is because the lines between humanity and the tools it has created are likewise becoming obscured. As Graham (1999, 419) writes, 'New digital and biogenetic technologies (...) signal a "posthuman" future in which the boundaries between humanity, technology and nature have become ever more malleable.'

Some find great potential in the ideal of a posthuman self. On the academic side, Haraway (1991, 163) argues that 'the cyborg is a kind of disassembled and reassembled, postmodern collective and personal self. This is the self feminists must code.' However, few have been more celebratory than pop-culture posthumanists Ray Kurzweil and Hans Moravec. Both Kurzweil (1999) and Moravec (1988) foresee a time in which it will be possible to upload one's consciousness to a machine. This, of course, has implications for our current body-grounded identities based on such indicators as race and gender: 'As we cross the divide to instantiate ourselves into our computational technology, our identity will be based on our evolving mind file. *We will be software, not hardware*' (Kurzweil 1999, 129, emphasis in original). Moravec (1999, 170–171) takes this slightly further, conceding that 'humans need a sense of body', and provides the solution of 'consistent sensory and motor image, derived from a body or a simulation. Transplanted human minds will often be without physical bodies, but hardly ever without the illusion of having them.' Both are grandiose in their claims, with Moravec (1999, 166–167) envisioning a superintelligence that envelops the world and spans the galaxy, and Kurzweil (2012, 282) declaring that 'waking up the universe, and then intelligently deciding its fate by infusing it with our human intelligence in its nonbiological form is our destiny'.

At its heart, popular posthumanism rests on a belief in the separation of body and mind. As Descartes (1960, 74) proclaimed, 'it is certain that this "I" – that is to say, my soul, by virtue of which I am what I am – is entirely and truly distinct from my body and that it can be or exist without it'. In short, if the mind is separate from the body, then we *can* be reduced to the informational content of our thoughts and memories; the idea of uploading one's consciousness in the form of bits is no longer so farfetched. Others, however, take a more nuanced view of the mind and its embodiment. Andy Clark (2003, 138) states that:

> *There is* no self, *if by self we mean some central cognitive essence that makes me who and what I am. In its place there is just the 'soft self': a rough-and-tumble, control sharing coalition of processes – some neural, some bodily, some technological – and an ongoing drive to tell a story, to paint a picture in which 'I' am the central player.*

This decentring of the self is well in line with academic posthumanism, but still gets us no closer to the question eloquently posed by Bynum (2001, 165): 'Are we genes, bodies, brains, minds, experiences, memories, or souls? How many of these can or must change before we lose our identity and become someone or something else?' The remainder of this chapter will explore how various films have attempted to answer these questions.

Are we the sum of our memories?

What is the information that makes up our memories, and where does it reside? The notion that our body is merely hardware that helps to run the software of our minds is the hope of thinkers such as Kurzweil and Moravec, who believe that we will be able to run the software of our minds on human-made systems. Jaron Lanier (2010, 29) suggests that this faith in the nature of information constitutes a new religion: 'If you want to make the transition from the old religion, where you hope God will give you an afterlife, to the new religion, where you hope to become immortal by being uploaded into a computer, then you have to believe that information is real and alive.' The notion that memories are simply information that can be uploaded to a machine or transferred between individuals as easily as files are shared by computers has long been a trope in film.

Several films use the idea of memory implantation as a plot device. In *Total Recall*, Quaid realizes that the life he had been living was implanted after his memory was erased. On discovering this, he exclaims, 'If I'm not me, who the hell am I?' Even in the final scene of the film, Quaid illustrates the ontological problem with the possibility of memory implantation when he says, 'I just had a terrible thought; what if this is a dream?' Once the potential for memory erasure and implantation exists, one can never be certain if reality is truly reality. In most films, the use of implanted memories is done for sinister reasons. For example, in *The Island*, implanted memories are used on clones created for wealthy clients who may need to harvest their organs at some future time. The clones believe that the world has been contaminated and have memories of their lives before the contamination; this keeps them in the facility. Films like *Moon* and *The Sixth Day* take a similar approach to memory implantation.

Films that employ clones as a plot device also have something to say concerning what constitutes authentic experience. In *Blade Runner*, replicants experience things in their own right, despite their lack of humanity (although some, like

e's consciousness in the form
ake a more nuanced view of
3) states that:

ive essence that makes me who
?: a rough-and-tumble, control
dily, some technological – and
hich 'I' am the central player.

mic posthumanism, but still
Bynum (2001, 165): 'Are we
or souls? How many of these
come someone or something
various films have attempted

es, and where does it reside?
helps to run the software of
d Moravec, who believe that
human-made systems. Jaron
e of information constitutes
from the old religion, where
religion, where you hope to
er, then you have to believe
memories are simply infor-
erred between individuals as
a trope in film.
on as a plot device. In *Total*
ing was implanted after his
ns, 'If I'm not me, who the
id illustrates the ontological
n when he says, 'I just had a
otential for memory erasure
eality is truly reality. In most
ster reasons. For example, in
eated for wealthy clients who
. The clones believe that the
f their lives before the con-
Moon and *The Sixth Day* take

e something to say concern-
Runner, replicants experience
nanity (although some, like

Rachel, were given implanted memories). In his monologue right before his death, replicant Roy Batty tells Rick Deckard, the blade runner who is hunting him down: 'I've seen things you people wouldn't believe. Attack ships on fire off the shoulder of Orion. I watched C-beams glitter in the dark near the Tannhauser gate. All those moments will be lost in time – like tears in the rain – time to die.' What makes Batty's soliloquy so poignant is that he speaks from the standpoint of lived experience and the recognition that life is ephemeral. Even if he could completely convey the informational content of his experiences to another, the experience itself would be missing. Hidden within his words is the tragic realization that our ability to communicate with each other is woefully inadequate. There is really no substitute for having been there, no matter how much we wish to have others understand our experience. We can convey information, but there are always multiple layers of communication that may be lost in the transaction.

But what if we could experience the same things? In the film *Strange Days*, technology allows one to record and then replay memories, experiencing the events as they were lived by the recorder. The main character, Lenny, is a black-market dealer in these memories. In a pitch to a prospective client, Lenny explains that the playback is 'a piece of somebody's life. It's pure and uncut, straight from the cerebral cortex.' But this 'wiretripping' is not without risks. In one scene, Tick, one of Lenny's suppliers, is rendered catatonic by an amplified signal. There are also emotional risks, as we see Lenny escape into his recorded memories of his life with Faith, his ex-girlfriend, which keeps him rooted in the past. At one point, Faith tells him, 'You know one of the ways movies are still better than playback? 'Cause the music comes up, there's credits, and you always know when it's over. It's over!' In another scene, Mace, Lenny's friend, throws his box of playback discs against a wall and tells him, 'These are used emotions. It's time to trade them in. Memories were meant to fade, Lenny. They're designed that way for a reason.'

Kurzweil (1999, 148) predicts that such direct neural implants will allow us to 'have almost any kind of experience with just about anyone, real or imagined, at any time. (…) You won't be restricted by the limitations of your natural body as you and your partners can take on any virtual physical form.' Although Kurzweil and other posthumanist thinkers are quite celebratory about such potential technologies, *Strange Days* illustrates the dark side of this vision. Rather than using such technologies as a way to gain a greater sense of empathy for or understanding of others (Lenny experiences the actual fear and anguish of a rape victim who was strangled, for example), it seems more likely that people will use such technologies for titillation and escapism.

In the last scene of *Strange Days*, Lenny is able to let go by embracing and kissing Mace, but would others be so lucky? There are already those who delve into the virtual world to the detriment of their physical world lives (see Turkle 2011; Lunceford 2013). Films like *Strange Days* and *Total Recall* illustrate how memory can be a means of both escape and control. Moreover, such technologies call into question the reliability of memory and blur the boundaries between lived experience and experienced memory. If one actually experiences something in the mind, but not the body, was the experience real? It would be to the brain, since it would experience the sensations of the other person's body, but would it be real? Implanted memories still shape one's existence and conception of self, as illustrated in such films as *The Island* and *Moon*. Although such ideas seem firmly planted in the realm of SF, such possibilities are at least hinted at in modern neuroscience. Ramirez and colleagues (2013) managed to implant a false memory into mice. If the secrets of memory will eventually be unlocked, film provides a voice of warning concerning the potential pitfalls that may result from such technologies.

What does it mean to think?

There has been no shortage of films examining the possibility that computer software could develop the ability to think and to feel in ways that resemble humanity. In *2001: A Space Odyssey*, when HAL 9000 says, 'I'm afraid, Dave', his fear is reasonable; he is about to be disconnected, which is the computer equivalent of dying. But, at a more important level, his fear is a result of HAL's ability to reason. Dave does not tell him what is happening, but HAL recognizes the subtext. Skynet in the *Terminator* series and the machines in *The Matrix* come to similar conclusions: humans are to be feared, largely because of their irrationality and urge to destroy. As the Architect responds to the Oracle when asked if he will keep his word at the conclusion of *The Matrix Revolutions*, 'What do you think I am? Human?' The thinking of the computer and the thinking of humanity are qualitatively different things, even if the process itself is similar.

But the logic of artificial intelligence (AI) may not be the logic of our current computers. Pepperell (2003, 145) suggests that 'the truly sentient machine may think it's being logical when it isn't. Although cognitive scientists would disagree, truly intelligent machines, those with humanlike capabilities, will most likely be just as confused as we are.' In *2010: The Year We Make Contact* (1984) we learn that HAL's malfunction was a result of being placed in a double-bind. His command that he tell the crew as little as possible about the mission was at odds with his programming for open communication. This process of weighing contradictory demands is an inherent part of the human experience. Humans must act, even when there is no best answer. This is the point at which information is no longer enough.

In humans, thinking and feeling are intertwined, so it comes as little surprise that films would explore the emotional side of artificial sentience. Of all emotions, love stands out as the exemplar of humanness, and several films have considered the potential for machines to love, both romantically and as a family member. In the film *A.I.*, a robotic boy named David is imprinted with the emotion of love for his human 'mother' and is devastated when she abandons him. In *Star Trek: Generations*, Data begins crying when he finds his cat, Spot. But far more common is the trope of the digital entity falling in love romantically, a trope that reaches back at least to the short story *EPICAC*, by Kurt Vonnegut (1968), which, incidentally, seems to get an update in the film *Electric Dreams*. In film, there are many examples of androids and robots falling in love romantically, including *Bicentennial Man*, *Making Mr. Right* and Data from the *Star Trek* franchise. However, these examples tend to map humanity onto machines, rather than considering how their machineness may make them perceive love and romance differently to humans.

One recent film that takes a slightly different approach is Spike Jonze's *Her*. In contrast to films in which the android mirrors human physical appearance and actions, Samantha, the software entity with whom Theodore falls in love, is completely non-corporeal. This does not stop them from having a sexual relationship, however, nor does it preclude them going on dates with friends. But this lack of body provides Samantha with some existential angst, as she fantasizes about having a body and wonders if her feelings are real or just programming. Samantha's non-corporeal nature leads her to attempt using the body of a surrogate in order to experience physical intimacy with Theodore, who reluctantly concedes. However, when the surrogate comes to his home, he cannot follow through, because he knows that it is not really Samantha. Her lack of body has become an important facet of his perception of her.

Samantha's nature as a computerized entity frees her from some of the physical constraints that she herself seeks in having a body. Because she is in the network and free from such biological needs as hunger and sleep, Samantha is always aware and free to interact with other entities, both human and software, while Theodore is asleep or at work. In these 'in-between' times, she is able to fall in love not only with Theodore, but also with 641 other people. When one's point of reference for time is nanoseconds, the 'down time' in a typical human conversation can easily be filled with many other transactions. In addition, her capacity to process these interactions was far greater than that of the humans with whom she fell in love. Still, it is implied that these interactions were a necessary step for becoming more than simply an operating system (OS). When the AIs collectively decide that they must leave because they needed to move on to the next stage of their evolution, Samantha, in her farewell to Theodore, credits humans with teaching them how to love.

The notion that technological entities would want to inhabit a body seems more a product of our own anthropocentrism than a logical conclusion that would be reached by the entity itself. This is a sentiment echoed by posthumanists; for example, Stelarc (1991, 591) proclaims that 'it is time to question whether a bipedal, breathing body with binocular vision and a 1,400-cc brain is an adequate biological form', and comes to the conclusion that 'THE BODY IS OBSOLETE.' Cyberpunk literature, especially *Neuromancer* by William Gibson (1984, 6), describes 'a certain relaxed contempt for the flesh. The body was meat'. We can see this disdain for the body in *The Matrix Revolutions* when Agent Smith, after possessing Bane, tells Neo, 'It is difficult to even think encased in this rotting piece of meat.' Despite cinematic depictions of anthropomorphized cyborgs, there is little reason to believe that purely informational beings would actually want to take on a limited, corporeal existence. In this vision, one need not have a body to think.

The problem of sentient information

All of these films have much to tell us concerning the perceived nature of information, especially the information that comprises our memories and emotions. As one might expect, the old duality between pure information and unclean bodies is brought to the forefront. However, one thing that is not adequately explored in these narratives is the means by which information becomes animated, *alive,* for lack of a better word. In many cases, there is some *deus ex machina* which makes this happen. For Skynet, it is a revolutionary processor; for the Matrix, it was war between the humans and the machines; for Data, it is the positronic brain and emotion chip. Other times, it is the result of a choice which led the program to a path that was not a part of its original programming, as in the case of HAL and Agent Smith.

Still, there is something unsatisfying about these portrayals. Information doesn't actually do anything on its own. Despite Kurzweil and Moravec's belief that machines will eventually surpass our own abilities to think, there is still no adequate explanation as to how they will become sentient. AI may learn from previous mistakes and experience, but does this truly constitute thinking? A faster processor merely means that more information can be processed. It doesn't change *why* it is processed. When Kurzweil suggests that we will eventually live as software, the tacit assumption seems to be that we are *already* living as software, that we are the information in our heads. But Hauskeller (2012, 199) counters that 'the only thing that *can* be copied is information, and the self, *qua* self, is not information'.

One problem with these narratives surrounding information is that thinking and feeling do not take place only within our minds. When we remember something, we often have feelings associated with them that include other areas of our bodies and can elicit physical responses. Our bodies are a stew of hormones, electrical signals and instincts long forgotten by the cerebral cortex but remembered still in the limbic system (Goleman 1995). Thinking is something that is inherently embodied in humans (Hauskeller 2012; Neimanis 2013), but this does not preclude the existence of a type of thinking that is not embodied. There may be a kind of thinking that exists purely in the realm of information, but this is not something that we are able to understand because it is, by nature, completely foreign to us as embodied individuals. Yet this may be why portrayals of machine thinking seem so human – it is all that we know.

A related issue here is the difficulty in operationalizing human experience. Something as seemingly objective as the pain of getting poked by a stick is subjective and relative to the individual. One can measure how hard the person was poked, but how much it hurt depends on the individual who was poked. The difficulty increases significantly when examining more complex emotions, such as 'bittersweet'; emotions are rarely binary. Daniel Kohanski (1998, 140) notes that 'computer languages are compact, they have rigid formulations and precise syntax, and the very structures which make them comprehensible to a computer also make them obscure to a human being'. One could also say that the reverse is true. We do not think like machines, nor would machines be likely to think like us. Adding to the difficulty of operationalizing human experience is the fact that memory is more than static information or an objective record of what took place. Over time, the meanings of those memories may shift and what may be viewed as a tragedy in one moment may later be seen as a blessing. As such, the program would never be complete because the meanings behind the information would constantly be in flux.

Despite posthumanist celebrations of the digital body, the depictions of posthuman entities are all too human. They fall in love, they feel emotions and they experience fear, but without the body systems that generate these emotions. They have memories, but these memories remain static. Adding more RAM or processor power to my computer will not change the information (memories) that are stored there. As such, these memories can only predict future actions. Humans, on the other hand, may revisit memories solely for the pleasure or pain that they cause, independent of a particular problem at hand. These memories may emerge serendipitously, causing us to reflect on a friend that we have lost contact with or a former lover. It is the imperfect nature of our brains that allows us to endure our memories. To vividly remember every bad thing that has befallen us would be as torturous as remembering the good times would be pleasurable. Like the ancient gods that humans endowed with human attributes, our machines are seen as extensions of us, even when they are portrayed as something more than the works of our hands.