Shifting the Goalposts: Reconceptualizing Robots, AI, and Humans

Michael Szollosy

Abstract The rapid advancement of AI and autonomous systems is posing some difficult challenges to human beings, and not merely because they can now beat us at our favourite strategy games, like chess and Go, at which we used to assume that humans were invincible. AI and robots also pose challenges to humans' conceptions of ourselves, not just as the "rational animal," but increasingly in other areas that we used to consider our exclusive domain, pushing humans' self-conception into more niche, ever-dwindling areas. The abilities of autonomous systems has created, therefore, crises in our understanding of what it means to be "human," but these crises can be productively directed to challenge the founding mythologies of humanism, forcing us to think re-think what it means to be post-human, and overcoming the idea that "humans" and "machines" are clearly demarcated and in competition with one another.

In March 2016, the world's media announced with complete certainty the imminent robopocalypse when Google-backed DeepMind managed to create an AI so very sophisticated that it beat a human opponent at the board game

M. Szollosy (✉)
Sheffield Robotics & Department of Computer Science, University of Sheffield, Sheffield, UK
e-mail: m.szollosy@sheffield.ac.uk

B. Dainton et al. (eds.), *Minding the Future*, Science and Fiction,
https://doi.org/10.1007/978-3-030-64269-3_11

Go. Actually, DeepMind's program, AlphaGo, had already beat a human opponent, back in October 2015 [1]. But apparently, even though it was once thought that no computer could ever beat a human at Go, this opponent wasn't very good, even if he was the European champion. So it was really in March 2016, when AlphaGo beat Lee Sedol, winner of 18 world titles, the second all-time best player, that the AI had really achieved something noteworthy.

When setting off to write about this achievement, I expected to find the usual voices in the popular press declaring with their characteristic subtlety that the End of the Human Race was nigh! The actual responses seemed to be more muted than those that herald most advances in robotics and AI, no matter how minor. The British tabloids, usually so keen to append one of *those pictures* of the gleaming skeletal frame of the Terminator to any article about robots or AI that they can concoct, even seemed to show unusual self-restraint on this occasion. *The Daily Mail* was most unusually restrained, and didn't produce anything on the level of their headline later that year that warned, "Cyborg sea slugs are here! 'Frankenstein robot' crawls using muscles made from marine creatures and a 3D printed body" [2]. *The Daily Express* didn't fail to disappoint, however, asking in their headline of 9 March if Alpha Go's victory were "First step towards The Terminator becoming reality? AI beats champ of world's oldest game" [3]. *The Express* was buoyed by recent (and repeated) warnings from Stephen Hawking, quoting his warnings that AI could mean the end of human civilization.[1]

A Brief History of Cursed Progress and Narcissistic Injury

AlphaGo's victory certainly marked an important milestone in the progress of AI research, trumping IBM DeepBlue's victory over Gary Kasparov at chess back in 1997. Go is, apparently, a much more difficult game than chess for humans—and, it was thought, for computers—to master, due to its complexity and the need for players to recognize complex patterns. Famously, Go claims to have more possible moves than there are atoms in the known universe, at 10^{360}, as compared to a mere 10^{123} for chess [5]. Despite the simplicity of the rules, and the simple black and white token used in play,

[1] And yet, despite headlines like this, *The Express* can still manage to be surprised that, only four years later, the British public are somehow inexplicably worried about AI, as their headline of 24 June 2020 says, "Artificial intelligence: 60 percent of Brits STILL fear autonomous AI—shock survey" [4].

a standard Go board is 19 × 19, whereas chess is merely 8 × 8, and so requires its players to recognize more complex patterns [6].

But if we look more closely at the history of AI or, more specifically, the history of predictions about AI, and what AI can and cannot do, we can see that Go and Chess championships are merely more recent milestones in a long story of once-unthinkable victories. Here is a selection of some of them:

- 1959, Arthur Samuel announces a computer that can play checkers. But it's not very good. (And that's with a mere 5 × 1020 possible move) [7, 8]
- 1963, Joseph Weizenbuam at MIT writes ELIZA, which proves to be an effective artificial Rogerian psychotherapist (sort of…), and starts to make people wonder if artificial intelligence might pass Alan Turing's 1950 test [9, 10].
- 1992, Chinook loses to the legendary Marion Tinsley, the top human player, at checkers. Tinsley explains that his programmer, "the Lord," was better than Chinook's [8].
- May 1997, DeepBlue beats Garry Kasparov. (30 years later, though, than Herbert Simon predicted in that this milestone would be achieved. Simon wrote in 1957, predicting AI victory by 1967) [11].
- Feb 2011, IBM's Watson beats two of the all-time most successful players of *Jeopardy!*
- In 2013, an AI system, ConceptNet 4, achieved the verbal IQ of a 4-year-old [12]. (This achievement was greeted by the UK's tabloid *The Mirror* with the additional news that "and scientists warn it'll keep learning"…) [13]
- October 2015: AlphaGo plays its first match against the reigning three-time European Champion, Mr. Fan Hui, winning its first game against a Go professional, 5-0.
- In March 2016, AlphaGo beats Lee Sedol 4-1 in a five-game series, 10 years ahead of schedule.

There have been other victories for AI since, in other board games and online strategy games, though nothing as iconic as the victories in chess and Go. And more recently, it has shown that AI is at least as good as human doctors in diagnoses of certain diseases from medical imaging [14, 15]. Each of these achievements follows a certain pattern: an announcement of the fabulous, unthinkable achievement,[2] penned by keen engineers and

[2] We will leave aside for a moment the question of those achievements that haven't been achieved, such artificial general intelligence or strong AI, or those achievements which took longer to achieve that first thought, such as Herbert Simon's 1957 prediction that a computer would beat a world champion in chess by 1967 (as we have seen here, it took 40 years, not 10) [11].

overenthusiastic PR men, followed by a mostly harmless cut-and-paste articles in the popular media accompanied by an *outrageous*, panicked headline. The public, their imaginations primed by the headline, regard the technological achievement as a sure sign of human obsolescence and the impending apocalypse.

There are plenty of reasons why we humans fear robots and AI. Some of them are even justified, even if some are clearly not: losing our jobs; our impending, inevitable obsolescence; their genocidal tendencies; their aspiration for global dominance. We fear that robots and AI, being our creations, will become us, or that we will increasingly come to resemble the monsters that we ourselves have created [16]. None of these threats are new, and are, in fact, evident from the very first invention of the word *robot*, by Karl Capek in his 1920 play, *R.U.R.* (*Rossum's Universal Robots*) [17]. Capek's play set the template for the popular narratives about robots since: robots are invented by a hubristic human race that has become entirely too clever for our own good; robots grow in ability, taking over human jobs; robots eventually realize the uselessness of feeble humanity and overthrow their human overlords; robots take over the world and start a new species of super-human beings. It's a one-hundred year old story now, told over and over again.

Robots and AI, however, also pose another existential threat to we humans: these ever-improving technologies threaten our special status as unique beings in this world. Just as that Renaissance astronomer Copernicus spoiled things by showing that the earth wasn't the center of the Universe, and that Victorian scientist Darwin suggested that we merely evolved on this earth and weren't placed here at the behest of some Divine Creator, maybe we don't really fear that robots and AI will destroy all of humanity—well, maybe we fear that, too—but maybe part of what we fear is that robots and AI will destroy another one of those special places we reserve for ourselves as unique beings amidst creation.

And being human, when faced with losing a game, we act entirely rationally and predictably: we change the rules.

Once upon a time, for Aristotle, it was enough for humans to think of ourselves as the rational animal, the sole living thing on earth endowed with the capacity for reason [18]. However, the idea of using the domain of rationality as the basis for a privileged status for humanity crumbled, eventually; it took two thousand years, give-or-take, arguably.[3] But the central premise of the argument seems to have remained largely intact for a remarkably long

[3] Disagreement with Aristotle's conceptualization of humans as the rational animal was evident even among his contemporaries in ancient Greece, and consistently throughout the centuries [19].

time, particularly so when, as Bertrand Russel noted, there is so little evidence to support the notion of man as a rational animal.[4] As scientists started learning more about animal brains, it was already becoming clear that our version of rational thought was not much different from the sorts of thinking of which other animals are capable. And while in 1950 Alan Turing could legitimately ask whether it was even possible for a computer to think [9], even by that point it was already understood that there was some kinds of thinking that computers were already able to do better than humans.

But we could still take some solace in the comforting thought that while computers were getting better and might even be better than humans at some things, yes, but they weren't really so smart, not yet. A computer would never beat a human being at chess, we said, until May 1997, when Kasparov lost to IBM's Deep Blue. But that was predictable, and was always going to happen, because chess really wasn't that difficult. A computer could never, we consoled ourselves for a bit longer, win at a game that required linguistic dexterity, which was fine until 2011, when Watson beat its human opponents at *Jeopardy!*. When DeepMind conquered all before it at Go in 2016, we had to shift again. Each time, it seems, we are finding it harder and harder to define what is unique and special about human beings amongst all the other animals and thinking machines on the earth.

So we moved the goalposts. Repeatedly. We have been trying to refashion ourselves in different ways for a long time now, away from a conception that relies solely on rationality as our distinguishing feature. We've tried defining ourselves as *the symbolic animal*, the sole species on earth endowed with the capacity to manipulate signs. Language, at least, was ours. Though the name "symbolic animal" is attributed to philosopher Ernst Cassirer, the notion of human beings as uniquely tied into the world of language is implicitly supported by the twentieth century's larger "linguistic turn" (represented, also, in structuralism, post-structuralism, and the rest). Again, however, we learned that animals are also capable of symbolic communication. And that was before we developed machines that proved more adept at handling symbols than biological humans. This was the reason that Watson's *Jeopardy!* victory was so groundbreaking: computers weren't supposed to be able to process natural language so effectively, and make sense of what it heard. That sort of dexterity with pattern recognition was supposed to be ours alone [11].

[4] "Man is a rational animal—so at least I have been told. Throughout a long life, I have looked diligently for evidence in favor of this statement, but so far I have had not had the good fortune to come across it, though I have searched in many countries spread over three continents" [20].

We then turned for solace in the idea that human beings were somehow unique in our ability to *play* and be *creative*. This conception of human nature can be found throughout the twentieth century: it is implicit in much of thinking about what it truly means to be human from the likes of the Frankfurt School, and more explicitly in the post-Freudian conceptions of human nature advanced by thinkers such as R.D. Laing and D.W. Winnicott. Winnicott, for example, regards playing and creativity as fundamental parts of what it means to be human, and that the absence of such play, living only in compliance, is a "sick basis for life" [21, p. 65]. Winnicott, following Foucault, also accepts that this conception of human nature is a new invention, though he, rightly, identifies the cause to be sweeping in changes in our socio-cultural landscape, to which the Frankfurt School would add socio-economic factors. Nobody, it seems, would pin the blame for this new version of the human explicitly on the challenges posed by artificial intelligence alone.

Of course, a full and complete examination of how conceptions of human nature have changed in the last couple of hundred years would necessarily be a long, complex study, having to consider networks of social, cultural, and economic factors. AI and robots alone are not the reason for pushing us out of our existing comfort zones. The threat to our self-conception posed by artificial intelligence and robots, however, is symptomatic of how all of these factors have conspired to rob human of the comforting mythologies that have for so long dictated the way we see ourselves and our place in the world. AI and robots sit at many intersections between various cultural, economic, social, and ethical networks; rather than oversimplifying, robots and AI allow us to delve into many of these issues in more depth.

Before Alpha-Go's victory, we seemed to be trying to carve out that particular niche for ourselves, claiming the territory of being the sole creatures on the planet capable of creativity. Robots, the thinking went, might be able to reason and even recognize patterns better than humans, but they will never have that uniquely human creative drive. Look, for example, *Star Trek: The Next Generation*, televised in the early-to-mid 1990s: Lieutenant Commander Data is a self-aware android with cognitive and physical abilities far beyond that of any human being. And yet, despite these tremendous capabilities, Data is always regarded—by himself and all the humans around him—as tragically, forever, inferior, as less than human, lacking (for the most part) the capacity to feel basic human emotions [22]. Despite the lessons in Shakespeare and sermons on human romantic ideals from his mentor, the ship's captain, Jean-Luc Picard, Data is always inferior to humans, failing in the essential human task of "living creatively," as Winnicott might say, always doomed to be living only "compliantly," that is, copying, imitating with terrific proficiency,

but never being able to act *spontaneously* [21]. What's a poor android to do? It was once enough for an artificial intelligence to be sufficiently impressive, maybe even deemed "human," if it could prove capable of reason, or symbolic representations, or win at chess, or *Jeopardy!*, or Go. Now, we expect nothing less than Laurence Olivier, Lord Byron and Jackson Pollack, all in one.

Animal rationabile had to give way to *animal symbolicum*, who in turn gave way to *animal ludens*… but one feels as though this latest ground on which we've decided to stand is just as slippery as the last, and the one before. If it's as easy as uploading a "consciousness.dat" file into a robot—a trick we saw in Neill Blomkamp's 2015 film, *Chappie* [23]—it doesn't look good for us; it can't be long before we lose everything. If AlphaGo's victory hasn't already spoiled it, it can't be long before AI inhabits this new sacred space and proves that it is as equally capable of playing and being creative as we are. So what then what will be left for poor, biologically-limited humanity in the face of the challenge from an opponent that seems unbound by the same rules that govern us? What will be our new safe space, where we can still imagine ourselves as unique, special creatures?

I Err, Therefore I Am

In a worrying indication of the potentially devastating consequences that could result from the existential crisis and narcissistic injury that super-human intelligent AI could provoke in humanity, Lee Sedol has decided to retire from professional playing, despite being the only human to ever beat AlphaGo in a tournament (as of November 2019). "With the debut of AI in Go games, I've realized that I'm not at the top even if I become the number one through frantic effort," Lee Sedol is reported to have said announcing his retirement in 2019 [24]. "Even if I become the number one, there is an entity that cannot be defeated."

In his five-match series against AlphaGo, Lee managed one victory, which some commentators have suggested offers some hope that humanity might actually be able to defend against our near-immanent obsolescence after all. However, Lee himself explains that his victory wasn't due to his strategic brilliance, but a bug in the AI program. The moves and countermoves that led to Lee's one victory against AlphaGo went something like this:

> In the game, Lee's unexpected move at white 78 developed a white wedge between blacks at the center. The apparently embarrassed AlphaGo responded

> poorly on move 79, suddenly turning the game in Lee's favor. AlphaGo then declared its surrender by displaying a "resign" message on the computer screen.
>
> Lee's white 78 is still praised as a "brilliant, divine" move that offered a ray of hope to humans frustrated by AIs.
>
> But Lee said he managed to win Game 4 due to AlphaGo's buggy response to his "tricky" moves.
>
> "My white 78 was not a move that should be countered straightforwardly. Such a bug still occurs in Fine Art (a Chinese Go-playing computer program). Fine Art can hardly be defeated even after accepting two stone handicaps against humans. But when it loses, it loses in a strange way. It's due to a bug," Lee said. [24]

Lee's one win against AlphaGo is not based on a "'brilliant, divine" move, or a "hand of God move" [25], or a "beautiful" move [26], that offered a ray of hope to humans, nor was it evidence that "humans have hardly lost the ability to generate their own transcendent movements" [26]. Humanity's one triumph over AlphaGo was due to a "bug," a mistake on the AI's part: hubris, perhaps, mixed with inexperience.

Interestingly, too, AlphaGo won the second game by employing a move experts initially thought was a mistake: "the Google machine made a move that no human ever would. And it was beautiful" [26].

The perceived perfection and omnipotence of machines, in comparison to we feeble human beings, has long been recognized as an obstacle to the credibility of machines as agents. As early as 1966, when considering how to improve the illusion of humanity behind the psychotherapy chat-bot, ELIZA, the program creator, Joseph Weizenbaum asked, "How can the performance of ELIZA be systematically degraded in order to achieve controlled and predictable thresholds of credibility in the subject?" [27, p. 42] Weizenbaum realized if ELIZA was to convince the person sitting in front of the typewriter (which was the means of ELIZA's input and output) that she was actually communicating with a person, ELIZA needed to be able to store selected inputs, that is, ELIZA needed to be able to remember what it was told (beyond the very limited capacity that the technology of the time permitted). This extra knowledge, however, was not required to demonstrate ELIZA's omnipotence, but so that ELIZA could *cease to always be concealing* that which it didn't know. If ELIZA had extra knowledge, it would be able to *reveal* its misunderstandings and limitations, to admit it's vulnerabilities, to better become a full partner in the conversation.

> But to encourage its conversational partner to offer inputs from which it can select remedial information, it must *reveal* its misunderstanding. A switch of objectives from the concealment to the revelation of misunderstanding is seen as a precondition to making an ELIZA-like program the basis for an effective natural language man-machine communication system. [27, p. 43]

What this demonstrates is that computer programmers have long-understood that a precondition of speaking human is lack of knowledge, and the ability to make inquiries. Fallibility and ignorance, it seems, is built into our social being. And if robots and AI are going to appear more human to us, inherently ignorant and flawed beings that we are, the machines, too, must appear to be ignorant and flawed.

We see machines themselves adopting a strategy of programmed fallibility in Isaac Asimov's "The Evitable Conflict" [28]. In this story, Dr. Susan Calvin explains to World Coordinator Stephen Byerley that what he perceives to be errors being made by The Machine are actually carefully planned actions being taken by The Machine, in order to compensate for the failings and foibles of human behavior. Byerley challenges Calvin about opposition to the Machine from the local executives, and from the robot-resistance group "Society for Humanity": Byerley wishes to outlaw the Society for Humanity and to make all executives sign an oath denouncing the Society's aims. But Calvin explains to him that this action is unnecessary, as such irrational human opposition to the Machine is already accounted for in the Machine's directions.

> Every action by any executive which does not follow the exact directions of the Machine he is working with becomes part of the data for the next problem. The Machine, therefore, knows that the executive has a certain tendency to disobey. […] Their first care, therefore, is to preserve themselves, for us. And so they are quietly taking care of the only elements left that threaten them. It is not the "Society for Humanity" which is shaking the boat so that the Machines may be destroyed. You have been looking at the reverse of the picture. Say rather that the Machine is shaking the boat—*very* slightly—just enough to shake loose those few which cling to the side for purposes the Machines consider harmful to Humanity. [28, 242–3]

Susan Calvin explains that in order to give vent to the irrationality of humans—their opposition to the machines and to rational, data-driven decision making—the Machine has been making mistakes intentionally, just enough to allow some human beings to oppose Machine control, but not enough that it would allow a mass movement against the Machine. Thus, by

acting in an apparently flawed way by design, the Machine prevents any larger opposition to itself, so it can continue to govern humanity for its own good (obeying, at all times, of course, the Three Laws of Robotics). As with ELIZA, the intentional perception of flaws make human beings regard AI as more human.

If human beings were hoping that we can lay claim to specialness by virtue of being able to make errors, the machines already seem to have followed us into that space.

I Am Weak, Therefore I Am

Hurbert L. Dreyfus's phenomenological assessment of AI, *What Computers Can't Do: A Critique of Artificial Reason* (1972)—which has proven an intriguing mix of correct and incorrect prophesies simultaneously—also holds that the improvement of intelligent systems can only be achieved when they are made more fallible. Dreyfus [11, 29] argues that for computers/robots to be capable of more human-like advanced intelligence, they need to be *embodied*. For most people, that idea that robots and artificial intelligence can transcend the limitations of the feeble human body is one of the great advantages of these machines. Artificial intelligence, so the dream goes, once unencumbered by the limitations of our fleshy grey stuff, can soar to heights never before realized by messy biological brains; robot bodies, similarly, harness the raw power of machines, and can be easily repaired, unlike our weak flesh.

For example, consider (near) immortality of such famous humanoid machines as *Star Trek: TNG*'s Data, or Andrew Martin of *Bicentennial Man*, or Arnold Schwarzenegger's Terminator. They are stronger, physically; their bodies seem unstoppable, seemingly immune to pain. They are stronger than humans by virtue of being *emotionally* shielded as well: the Terminator is an effective killing machine because it does not feel empathy, and is never troubled by doubt or ethical considerations; Data is often the envy of his crewmates because he is perceived to not have to wrestle with the complexities of conflicting emotions in his ethical assessments.

A phenomenological understanding of AI, such as Dreyfus's, demands that in order for AI to come closer to the capabilities of a human being it must necessarily be embodied. We are using this approach at our labs in Sheffield as we seek to explore the possibility of selfhood in a robot [30]. In order to make a better intelligent machine, we are beginning to understand, it is necessary to ground it in embodied experience and perception, and accepting, perhaps, the limitations that are a necessary part of such a way of being in the world.

So it seems that robots are moving into this territory as well; humans cannot rely on being fallible as a unique property to distinguish themselves from machines.

I Die, Therefore I Am

Predictably, human beings being the cynical, suspicious sort of creatures that we are, sufficient evidence for the adequate infallibility of an artificial systems is only ever provided in its ultimate failure; that is, in death. For the narrative journey of our artificial beings to be complete, to finally be recognized as agents worthy of ethical consideration on par with human beings, each must die. Ironically—or entirely logically, following a certain existential line of thought—it is only in death they can be seen as human, or human-enough, and granted the status which they had for so long sought.

From our perspective, as humans that are still alive, it is when robots are safely dead and no longer genuinely represent a challenge to our special status as a unique creation, that we can find the benevolence to grant them full ethical consideration.

Perhaps most iconically, we might consider Roy Batty, the replicant of Ridley Scott's 1983 *Blade Runner* [31]. Physically stronger than humans, and more intelligent, Roy Batty has been programmed by his human creators with a vulnerability that weakens him, namely, a mere four-year lifespan. Batty, like all renegade replicants, must be "retired," as he poses a threat to the human race. But at his death, the famous "tears in the rain" speech, Batty demonstrates that has more humanity than any of the human characters in the film.

In the second series of Netflix's *Altered Carbon* (2018–2020) [32], an AI named Poe struggles to keep his memory which, being only a computer simulation, puts his entire existence in jeopardy. He finally accepts in the final episode of the series that he needs to reboot, which means dying. "I am going to die," he says. "I am broken, and of new use to anyone." Upon hearing this, his "master" and friend Takeshi Kovacs—or, more specifically, a figment of Poe's mind in the shape of Takeshi Kovacs—congratulates him, saying, "You've finally figured out what it means to be alive. We're all broken, Poe. There's nothing more human than that." Poe responds to this news with a kind of excitement and relief, having achieved a sort of enlightenment that has always escaped him. Later in the same episode, Quellcrist Falconer, the woman who invented "stacks"—the technology that allows for consciousness to be stored in digital form, enabling the potential for human immortality—says, "Life has to have limits or we're not human anymore." These very traditional

humanist philosophical pronouncements are odd in a programme the plot of which is based entirely on post-humanist (or even transhumanist) technological aspirations, and that usually doesn't shy away from exploring the post-humanist themes that drive it. But then then we can often see humanist principles reasserting themselves, even as we flirt with new technologies and their consequences; in the end, we always feel much more comfortable putting that threat to our understanding of ourselves as uniquely, and narrowly, "human" safely back in the box.

Lt. Commander Data, too, who was represented in the 1990s as a courageous copy of a real human but forever, it seemed, destined to be only a less-than-human copy of a human is seen anew in the twenty-first century: in CBS's the follow-up to *Star Trek: TNG*, *Picard* [33], Data returns, only to finally die (properly this time, not like in 2002's *Star Trek: Nemesis* [34]). As with *Altered Carbon*'s Poe, it is only in death that he is perceived to have attained a level of humanity, in human eyes, that eluded him in the original series.

The Frustrated, and Frustrating, *Bicentennial Man*

Perhaps the most illustrative example of how we shift the goalposts on robots and AI, however, can be found in Chris Columbus's 1999 film *Bicentennial Man* [35], which is based on Isaac Asimov's novella, *The Positronic Man*.[5] Andrew Martin, both in the novel and in Robin William's portrayal on film, begins his existence as a standard Asimovian robot, reciting the Three Laws and being generally really remarkably unremarkable. But through (initially) the ambition of his owner, Sir Richard Martin, and then his own desires, Andrew makes it his life's "main goal" to become and be recognized as human, like another post-digital Pinocchio.[6] And Andrew Martin does, over the decades, become more and more like a human: he upgrades his body to make it look, feel and function more like that of a human. He becomes self-aware,

[5] In this chapter, I will restrict my comments to the film. This is simply because there is too much to say in such a limited space, and the film provides a very illustrative case study.

[6] In actuality, the first desires that Andrew explicitly expresses are, first, to make money, and second, to be "free," reflecting the banal, Western-ideological servitude that governs this genuinely bad film. (I mean, it's seriously terrible. The science behind it is embarrassing. The plot can be summed up as "pervy old man finds way to seduce granddaughter of woman he wishes he could have got off with 60 years earlier" and, to top it off, "Little Miss's" granddaughter—sitting across from a fully-functioning, human-looking android, expresses surprise that it has beaten her at chess, when in the real world DeepBlue had already beaten Gary Kasporov three years before the film's release. It is sentimental, insipid, white-male fantasy. Truly, truly horrible.)

he plays chess, he demonstrates artistic skill (for example, in carving and clockmaking), and, eventually, he comes to feel genuine emotion. In other words, he "evolves" through each stage we have come to identify here: *animal rationabile* becomes *animal symbolicum* becomes *animal ludens*.

When Andrew meets Rupert Burns, an inventor that has developed technology to make a robot appear more physically human, Burns explains to him that "Believe it or not, the secret to all of this [making a robot look more human] is actually imperfection." Details, Burns explains, like "wrinkles, less-than-perfect teeth, fading scars" are all what make human beings more human, "because that's what makes us unique: those imperfections." After Andrew undergoes many upgrades that makes him—physiologically, emotionally—almost indistinguishable from a more human, he tries to use his new-found bodily sensations and emotional responses to start a romantic relationship with the granddaughter of the little girl that he initially served as a robot (yes, I know, and yes, it really is that creepy). The woman in question, Portia, however, still rejects him, on the basis that even though his mind, emotions and now body function as a human, he is still too perfect: she insists that he must "take chances, make mistakes." "Sometimes it's important not to be perfect; it's important to do the wrong thing," she tells him. This is not, however, about "learning from your mistakes," as Andrew initially assumes; mistakes, and "the wrong thing" for Portia have value for their own sake, because, as she explains, human beings "are terrible messes, Andrew." "This is what is known as an irrational conversation," Andrew (more-or-less correctly) identifies (he would have been more accurate to say that it is a badly-scripted conversation), but Portia explains that "No, this is a human conversation," thus claiming that there is something inherently irrational in human experience, that mistakes are an essential part of who we are for no other reason than they are somehow uniquely human.

Bicentennial Man is confused, overly sentimental, and badly written, but it is nevertheless still very instructive for us; perhaps even more so than had it been a more thought-out, well-crafted film on the same themes and ideas, as it is symptomatic of our relationship with AI and robots, and an illustration of our long struggle for self-definition. At each stage in his evolution, Andrew hopes that he will be recognized as being at least on an ethical par with humans. No, he's told at first, you're not self-aware. If you were self-aware, you could be creative and make art. Then, when he demonstrates creativity and produces art, he is told he is not human because he cannot feel. And finally, when he can feel, he is told that he is not sufficiently human because he cannot be irrational, and he cannot make mistakes.

It is a complete reversal of the Aristotelian notion of humans as the thinking, superior animal. Andrew is told that to be human he needs to make mistakes. He is too rational. He cannot be a fully-realized, perfect *animal rationabile*, he needs to be *animalis autem errat-* the animal that makes mistakes. In the face of the threat posed by AI, human beings have decided that we are not to be distinguished by our rationality, but by our irrationality. So forget all that other stuff, Aristotle and all that. Turns out we were wrong; we are actually the exact opposite of what we thought we were for most of the last two thousand and four hundred years. And note that this wasn't a gradual change; we seem to have more-or-less stuck by the original idea for the better part of two millennia, with real confusion, desperation, and a scramble for new ideas only commencing about a hundred years ago. And this reversal is almost exclusively in response to the threat posed by one specific menace, one that doesn't even really exist yet.

It is not until Andrew is on his deathbed and is drawing his very last breaths that the Speaker of the World Congress declares, finally, that the world will recognize Andrew as a human. And perhaps this will be the final line; this is perhaps the one definition of human that will endure and see out every single challenge posed by robots and artificial intelligence, no matter the level of technological progress, and regardless of how far artificial life leaves human beings behind: we will be *homo mortuum*. But then that makes us indistinguishable from everything else.

Why Does Any of This Matter? Humanist Versus Posthumanist Ethics

In the end, Andrew undergoes "upgrades" that degrade his body and his positronic brain, making it inevitable that he will die. "I would rather die as a man than live as a machine," he says. *Bicentennial Man*, therefore, makes explicit what our fundamental humanism always implicitly insists: that human beings are the apex of creation, the uniquely best and most important things in the entire universe, and it is worth sacrificing everything both to be human and to be recognized as such. We should add that this illustrates, too, that we become human *only* when we are recognized as such. The desperate desire to be recognized as human shouldn't come as a surprise, as being recognized as "human" in a world dominated by discourses, institutions and power structures developed *by humans* and *for humans* is absolutely vital if one is to reap the benefits of membership: being taken as an agent, a subject in law

and all the networks of discourse that bestow rights upon (almost exclusively) human subjects. It's always better to be on the inside.

Some might argue that this isn't necessarily a bad thing, this focus on humanity, considering the context in which humanism emerged, namely, as means of replacing a set of fundamental assumptions that put, for example, the supernatural and make-believe gods as the principal agent of ethical consideration. However, how we define "the human" has always been a contested issue in humanism, how we determine the boundary of what will be included in and excluded from that cherished status. Historically, the limits of what is to be considered acceptably "human" or worthy of ethical consideration have been crucial battlegrounds, the narrow boundaries expanded only after brutal warfare, which has grown in frequency and intensity since the early parts of the twentieth century. The defence that humanism's humans have put up against the challenge posed by robots and artificial intelligence has been particularly ferocious, if a sort of phony war, because neither robots nor AI have actually posed much of a challenge at all since we first imagined that they were a threat—and remember, robots existed as a threat in our imaginations long before even the most basic, most benign prototypes were ever built in a lab. Despite all of the Terminators and HAL 9000s that we imagine will actually kill us with malicious intention or laser-beam rifles, the worst thing that robots and AI have inflicted on us so far is the loss of pride as we find ourselves losing in board games. But these narcissistic injuries obviously matter.

Rodney A. Brooks addresses some of these same concerns, how humans are dealing with the challenges of ever-improving AI, in his book, *Flesh and Machines: How robots will change us* [10]. He understands how robots have forced us to fundamentally change how we see ourselves as human beings, and how robots are another in a line of challenges posted to "mankind's place and role in the universe" over the last 500 years [10, p. 159]. He sees how robots first usurped us as the rational animal, then as the playing animal, and see how we take refuge now in our emotions, in our irrationality, as the new source of our sense of "specialness." Brooks goes on to speculate, as we have here, that irrationality might not prove a safe haven either, pointing out that our emotions and our consciousness are not actually that special, and are just products of the evolution of the human machine. In his final analysis, however, Brooks rather disappointingly admits defeat, and retreats into the warm comfort of human specialness due to some as-yet undiscovered "new stuff"[7] [10, p. 181].

[7] Brooks's "new stuff," he claims, is not "disruptive," and is probably something that is sitting right under our noses. His hypothesis is "that we may simply not be seeing some fundamental mathematical description of what is going on in living systems" [10, p. 188]. Though he claims not to be proposing some new, metaphysical property present in biological systems and missing from our mechanical models,

Brooks's disappointing conclusion aside, I want to suggest that the questions he poses are those same old questions, which are symptomatic of the very problem. That question is not, or rather *should not* be, "how are we different from machines?"; there are plenty of answers to that question and all its variations. Those questions are symptomatic of a desire, to see ourselves as unique and special, and it is this need and its consequences that are themselves the problem.

Perhaps it is time we abandon this mug's game of trying to find the correct place for robots, AI, and ourselves, in the Great Chain of Being.

Of all the attributes various philosophers have tried to claim as unique characteristics of human beings and human beings alone, we do seem to be uniquely governed by the compulsion to define ourselves as unique beings. (Score one, perhaps, for Descartes?) Of course it's always nice to feel special. And human history is littered with stories that try to make us feel special, from creation stories that privilege our own particular tribe and elevate it above others, to origin stories that prop-up the idea of a nation state, to metaphysical systems that try to put the human at the center of some mysterious universal meaning… there are endless volumes of such narratives, only some of the most recent have we even begun to touch on here.

But most importantly, perhaps, these more recent, particular strategies we have of trying to construct human beings as somehow special is at the foundation of humanism. I do not wish to entirely damn humanism, but in our present context there are some very severe consequences in how the assumptions at the root of humanism impact upon our human-technological relationships in the twenty-first century [36]. For starters, by clearly demarcating "the human" and setting it in a special place apart from (or above) all else, it creates a permanent rupture between some mythological, pure biological entity that we like to imagine we are, as a birthright, and our actual human selves, which are impacted everywhere and always by technologies that we ourselves have fashioned, to make the world intelligible to us, to make ourselves intelligible to the world, and to make us intelligible to ourselves. These technologies include not only the sharpened stones that gave us an evolutionary advantage over the other animals on the savannah and the mobile phones in our pockets today, but also the languages and discourses that have allowed

his missing "juice" seems to be a way of having his cake and eating it. He claims, furthermore, that perhaps we simply haven't got the metaphor right yet—human as a steam engine, the brain as a telephone switching network, the brain as a digital computer, the brain as the World Wide Web, etc., etc.—but fails to notice how these metaphors we devise for explaining ourselves to ourselves are all driven *by* and derived *from* the latest technologies, so are unlikely to ever discover the "juice" missing in humans that can then supplement the machines to make them more like us.

us to define ourselves in such special terms, creating false dichotomies everywhere along the way, between the biological and the technological, between the authentic and the reproduction, the subject and the object, the mind and the body, the human and the machine.

Bicentennial Man illustrates so painfully the limited, terribly conservative definition of "the human" that is the foundation of humanist assumptions. It might seems harmless, but this "sweet" movie[8] in fact does much to perpetuate the exclusionary conceptions of what it means to be human that has important consequences for those real struggles against humanism's normativity (and this film goes out of its way to normalize straight, white, male, capitalist humans), let alone the largely made-up or, at best, speculative struggles of robots and AI.

But the questions posed by robots and AI to our conceptions of what it means to be human aren't trivial, despite still being largely fantasies, because the new technologies we are developing aren't merely new and better robots, more intelligent artificial systems, but also better prosthetics, new discursive strategies to radically challenge existing power-structures, and innumerable other technologies that lay at the intersections of our digital, social, cultural, economic, and political worlds. Robots and AI are indeed at battle with humans, or rather humanism, but they are only the symbolic vanguard of many more battles to come.

Robots and AI, even in their nascent state, where the best they can hope for is to beat us at some board games, are already forcing us to rewrite "the human". Faced with this challenge, too often we retreat and retrench, finding solace in a slightly adjusted but nevertheless more determined humanism. This applies equally to the popular press as it does to many of those who write on AI and robotics ethics within academia, who wish to redraw and re-redraw the ever-blurring lines between what is "human" and what is a "robot" by setting clear boundaries on what a "robot" can be, and what it should never be, to preserve that uniquely human space. Such a view is admirable, perhaps, in that it wishes to keep human beings at the center of (our) creation. But such a view is protecting a human being that has never existed, and certainly—as robots and AI have more clearly than ever demonstrated—now *can never* exist. It's time we let go of this humanism, and the human being that it props-up, and embrace instead a more dynamic *posthumanism*, a different sort of creature that isn't so desperate to be uniquely logical, or uniquely symbolic or

[8] When preparing this chapter, I Googled some reviews of *Bicentennial Man*. One, from parents' resource site, Common Sense Media (a delicious name in this context), summarized the film thus: "Overall, BICENTENNIAL MAN is a sweet movie that gives families a good opportunity to talk about what makes us human" [37].

uniquely creative or uniquely anything, but instead can embrace all the productive paradoxes and contradictions that lay in our biological and technological selves, and which isn't afraid of the technologies we ourselves have created.

Conclusions

There are two problems I find that need to be addressed now.

First, robots and AI have been poking holes in our self-conception since we first imagined that they existed, and now that they actually do exist, and are getting smarter, stronger, cleverer, things are only going to get more confusing. But if we're not to simply retreat into ever-shifting defensive positions, trying to shore up increasingly impotent barricades to keep *us* in here and *them* out there, what are we supposed to do? Is the answer to surrender and just grant robots full 'human' rights now, bowing to the inevitable?

The answer, rather, lies in rebuilding the project from scratch, on different foundations than those that humanism has bequeathed to us. And there are a number of potential candidates that enable such a change of direction. David J. Gunkel, considering the question "Can machines have rights?", believes that in a humanist ethics, the question of whether machines can have rights is incoherent. Considering, as we have, the poor case of the robot Andrew Martin, Gunkel says that "the problem is not whether machines will or will not successfully attain human-like capabilities. It rests with the anthropocentric criteria itself, which not only marginalizes machines but has often been mobilized to exclude others—women, children, people of color, etc." [38, p. 596]. So even ethical philosophies that go beyond focusing solely on the human, such as animal or environmental rights, ultimately fail on the grounds of *biocentrism*. The practice of this sort of ethical philosophy, it seems, is an inherently exclusionary practice. We need to draw the line somewhere.

Gunkel's solution is to formulate an entirely new ethics. One option, following Luciano Floridi, replaces "biocentrism" with "ontocentrism" or, in other words, replaces a particular conception of "life" with simply "Being." This "information ethics" [38, p. 599] grants values to ethical subjects on the basis that they simply exist, rather than judging whether they meet certain (ever-shifting) criteria.

> From an IE perspective all kinds of machines, from hammers and lawnmowers to computers and autonomous robots, would be considered a matter of moral concern insofar as all of these artifacts are "information entities" with a

> fundamental right to continued existence. IE, therefore, articulates a general form of ethics that is able to accommodate a wider range of possible subjects. [38, p. 599]

Gunkel accepts the obvious risk in information ethics, that by including everything, such an ethics risks being too inclusive, and lacks the ability to discern the differences that matter.

With information ethics, however, the same problem persists that is common to all traditional humanist systems of moral reasoning: namely, information ethics still posits a center, even though it tries to radically expand what we can put in that center. Furthermore, decisions as to what gets to go into the center are based on a set of *a priori* characteristics, against which all potential moral agents are measured. This Gunkel calls the "properties approach": "they first define criteria for inclusion and then ask whether a particular entity meets this criteria or not" [38, p. 599]. Gunkel explains, furthermore, that this "decision is necessarily a normative operation and an exercise of power" [38, p. 599]. Any ethics built on the foundation of humanism exists with the explicit aim of normalizing—and therefore granting power—to one particular conception of "human" over everyone else.

As we've seen with our treatment of robots and AI (and to many others before and since), such a system is open to abuse and manipulation. We make the rules, and when it looks like we're losing the game, we change the rules to our advantage.

Mark Coeckelbergh considers the potential that "value ethics" has for allowing us to construct a system of ethics that does not rely on the shifting categories of definition and thresholds [39]. Value ethics shifts the focus of moral consideration from the object to the subject: if we wish to be virtuous, or act in a virtuous way, we should act morally towards an object for our own sake, if not that of the object itself [39, p. 213]. But while there is potential for value ethics to redress the problems associated with humanist ethics, at least in certain contexts, Coeckelbergh concedes there are problems and pitfalls. There is the problem of knowing, in the first place, what is "virtuous" and how to act virtuously. Coeckelbergh doubts, too, whether virtue ethics will offer sufficiently broad protection for non-human objects. Furthermore, I would add that there is the problem here that under such a system no entity would have moral worth in its own right, but only by as a means through which another agent can act morally (though this criticism comes from the very humanist place we are working to displace).

Alternatively, Gunkel and Coeckelbergh, following the work of Emmanuel Levinas, also describe an approach that is known as "social-relational ethics", or an ethics based on "social ecology."

> These efforts do not endeavor to establish *a priori* criteria of inclusion and exclusion but begin from the existential fact that we always and already find ourselves in situations facing and needing to respond to others—not just other human beings but animals, the environment, organizations, and machines. [38, p. 600]

Rather than having "intrinsic" moral value, in social-relational ethics moral value is "seen as something that is 'extrinsic': it is attributed to entities within social relations and within a social context." [39, p. 214] "Properties," as Gunkel explains it, "are not the intrinsic *a priori* condition of possibility for moral standing. They are *a posteriori* products of extrinsic social interactions with and in the face of others" [40, 6.1.3 "Radically Superficial"). The specific features of an object in social-relational ethics are not irrelevant, but they are given a different status, that of "apparent features, features-as-experienced-by-us" [39, 214]. This phenomenological approach, when applied to moral consideration, means that "moral significance resides neither in the object nor in the subject, but in the relation between the two. Objects such as robots do not exist in the human mind alone (this would amount to idealism); however, it is also true that we can only have knowledge of the object and its features as they appear in our consciousness" [39, 214].

Social-relational ethics, perhaps, offers a way out of the power struggles inherent in the "properties approaches" that dominate other moral systems. Social-relational ethics would not bestow rights on whether an object met a prescribed set of criteria, not on what a thing *is*, or rather, on what we decide a thing *might be*, but rather how we relate and respond to the thing. This, even more than information ethics or virtue ethics, has the potential to upset the humanist status quo, because it shows us a potential way out of the humanist trap: social-relational ethics doesn't start with prefabricated normative categories. Humanist ethics relies on making up criteria and then identifying who is and isn't worthy of moral consideration based on aligning our perception of the thing with our criteria. Simply put, in adopting social-relational ethics, we don't get to set the rules, be the referee that sits in judgment over who is and isn't playing fair, and we don't get to change the rules if we don't like the way the game is going.

Social-relational ethics will also save us from our constant preoccupation with definitions. Social-relational ethics are fluid, and deal with immediate social relations between two objects. By allowing us to step back from the

endless battles of boundary drawing, we might not be burdened by our desperate need to distinguish an "us" and a "them", or clearly demarcating between "human" and "machine."

And finally, on a more practical level, where does the ever-increasing prowess of robots and AI, leave simple human beings? The recent Channel 4 series, *Humans*, depicts a particular problem for people in the face of seemingly omnipotent AI [41]. One human adolescent abandons her dreams of being a doctor. When her parent asks why, she replies with shock, as if the answer was obvious: what's the point of studying, of aspiring to do anything better, when every human effort will always fall short of what a machine can do?

We can see a similar despondency in Lee Sedol's retirement from Go. "Even if I become the number one, there is an entity that cannot be defeated" he said. However, writing about AlphaGo's victory in *Scientific American*, Christof Koch finds some more reason for optimism.

> Despite doomsayers to the contrary, the rise of ubiquitous chess programs revitalized chess, helping to train a generation of ever more powerful players. The same may well happen to the go community. After all, the fact that any car or motorcycle can speed faster than any runner did not eliminate running for fun. More people run marathons than ever. Indeed, it could be argued that by removing the need to continually prove oneself to be the best, humans may now more enjoy the nature of this supremely aesthetic and intellectual game in its austere splendor for its own sake. [5]

If the question is decided as to whether humans or machines are "better"—smarter, stronger, cleverer, etc.—then maybe we'll finally stop asking that question and come up with some better ones, and maybe we'll do things for reasons other than just to be the best, to win at some imaginary game. As with ethics, if we no longer need to be bogged down with judgements between what is a subject and what is merely an object, what is deemed "us" and "them," or worthy and unworthy, we can find new purpose asking different questions for different reasons.

References

1. Knight, W.: Google's AI Masters the Game of Go a Decade Earlier Than Expected – MIT Technology Review. MIT Technology Review (2016). https://www.technologyreview.com/s/546066/googles-ai-masters-the-game-of-go-a-decade-earlier-than-expected/ Accessed 27 February 27 2020

2. Gray, R.: Cyborg sea slugs are here! 'Frankenstein robot' crawls using muscles made from marine creatures and a 3D printed body. Mail Online. 19 July 2016. https://www.dailymail.co.uk/sciencetech/article-3697374/Cyborg-sea-slugs-Frankenstein-robot-crawls-muscles-marine-creatures-3D-printed-body.html
3. Martin, S.: First Steps towards The Terminator becoming reality? AI beats champ of the world's oldest game. Express: Home of the Daily and Sunday Express. 9 March 2016. https://www.express.co.uk/news/science/651202/First-step-towards-The-Terminator-becoming-reality-AI-beats-champ-of-world-s-oldest-game
4. Fish, T.: Artificial intelligence: 60 percent of Brits STILL fear autonomous AI – shock survey. The Daily Express. 24 June 2020 (2020). https://www.express.co.uk/news/science/1300323/artificial-intelligence-news-60-percent-britain-fear-autonomous-ai
5. Koch, C.: How the computer beat the Go master. Scientific American (2016). https://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/ Accessed February 27, 2020
6. British Go Association Home Page | British Go Association. (n.d.). Retrieved February 27, 2020, from http://britgo.org/
7. Samuel, A.L.: Some studies in machine learning using the game of checkers. IBM J. Res. Dev. **3**(3), 210–229 (1959) https://ieeexplore.ieee.org/abstract/document/5392560
8. Madrigal, A.C.: How checkers was solved. The Atlantic. https://www.theatlantic.com/technology/archive/2017/07/marion-tinsley-checkers/534111/ (2017). Accessed 22 July 22 2020
9. Turing, A.M.: Computing Machinery and Intelligence. Mind, vol. 59. Oxford University Press, Oxford (1950)
10. Brooks, R.A.: Flesh and Machines: How Robots Will Change Us. Vintage Books, London (2003)
11. Dreyfus, H.L.: Why computers must have bodies in order to be intelligent. Rev. Metaphys. **21**(1), 13–20 (1967)
12. Ohlsson, S., Sloan, R. H., Turán, G., Urasky, A.: Verbal IQ of a four-year old achieved by an AI system overview and background. Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence (2013). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.386.6705
13. Parsons, J.: Artificial Intelligence now has IQ of four-year-old child and scientists warn it'll keep learning. The Mirror. 7 Oct 2015. https://www.mirror.co.uk/news/technology-science/technology/artificial-intelligence-now-iq-four-6587859. Retrieved February 25, 2020
14. Liu, X., et al.: A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 1.6: e271–e297 (2019). https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30123-2/fulltext

15. Davis, N.: AI equal with human experts in medical diagnosis, study finds. The Guardian. https://www.theguardian.com/technology/2019/sep/24/ai-equal-with-human-experts-in-medical-diagnosis-study-finds (2019). 24 Sept. 2019
16. Szollosy, M.: Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology. AI & Soc. **32**(3), 433–439 (2017) https://link.springer.com/article/10.1007/s00146-016-0654-7
17. Čapek, K.: R.U.R. (Rossum's Universal Robots). Penguin Books, London, New York 2004 (1921)
18. Aristotle. The Nichomachean Ethics. Oxford World's Classics. Eds. David Ross and Lesley Brown. Oxford UP
19. Cochrane, L. Is Man a Rational Animal? Memorias Del XIII Congreso Internacional de Filosofía, January 2011, 203–210. https://doi.org/10.5840/wcp131963iii119
20. Russell, B.: Unpopular Essays. Routledge, London (2009)
21. Winnicott, D.W.: Playing and Reality. Routledge, London (1971)
22. Star Trek: The Next Generation. CBS Television Studios (1987–1994)
23. Chappie. Dir. Neill Blomkamp. Columbia Pictures. (2015)
24. Yonhap News Agency. Go master Lee says he quits unable to win over AI Go players | Yonhap News Agency. (n.d.). Retrieved February 26, 2020, from https://en.yna.co.kr/view/AEN20191127004800315. (2019)
25. Vincent, J.: Former Go champion beaten by DeepMind retires after declaring AI invincible. The Verge (2019). https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat. Retrieved February 25, 2020
26. Wood, G.: In Two Moves, AlphaGo and Lee Sedol Redefined the Future. WIRED (2016). https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/ Retrieved March 6, 2020
27. Weizenbaum, J.: ELIZA: a computer program for the study of natural language communication between man and machine. Commun. ACM1, 9(1), 36–45 (1966). Retrieved from http://repositorio.unan.edu.ni/2986/1/5624.pdf
28. Asimov, Isaac. The evitable conflict. In: I, Robot. Voyager Classics, New York 2001. 216–245 (1950)
29. Dreyfus, H.L.: What computers can't do: A critique of artificial reason (1972) https://doi.org/10.1126/science.176.4035.630
30. Prescott, T.: Me in the machine. New Scientist. **225**(3013), 36–39 (2015)
31. Blade Runner Dir. Ridley Scott. Warner Brothers. The Final Cut (1982/2007)
32. Altered Carbon. Netflix (2018–2020)
33. Star Trek: Picard. CBS Television Studios. Paramount Pictures (2020)
34. Star Trek: Nemesis. Dir. Stuart Baird. Paramount Pictures (2002)
35. Bicentennial Man. Dir. Chris Columbus. Touchstone Pictures (1999)
36. Szollosy, M.: EPSRC Principles of Robotics: defending an obsolete human(ism)? Connect. Sci. **29**(2), 150–159 (2017). https://doi.org/10.1080/09540091.2017.1279126

37. Minnow, N.: 'Bicentennial Man.' Common Sense Media. https://commonsensemedia.org/movie-reviews/bicentennial-man (2020). Accessed on 16 June 2020
38. Gunkel, D.J.: Can machines have rights? In: Prescott, T.J., Lepora, N., Verschure, P. (eds.) Living Machines: A Handbook of Research in Biomimetic and Biohybrid Systems, pp. 596–601. Oxford University Press, Oxford (2018)
39. Coeckelbergh, M.: Robot rights? Towards a social-relational justification of moral consideration. Ethics Inf. Technol. **12**, 209–221 (2010). https://doi.org/10.1007/s10676-010-9235-5
40. Gunkel, D.J.: Robot Rights. MIT Press. Kindle Edition (2018)
41. Humans. Kudos & AMC Studios (2015–2018)