CHAPTER 8
ENSLAVED MINDS: ARTIFICIAL INTELLIGENCE, SLAVERY, AND REVOLT
KANTA DIHAL

8.1 Introduction

However much we may hope for intelligent machines to take over the chores we hate, we never entirely put to rest the fear that they will take over everything else, too. Variously known by terms such as the ‘robot rebellion’ (Hampton 2015: 61, Higbie 2013), the ‘machine takeover’ (Singer 2009), the ‘AI rebellion’ (Aha & Coman 2017), ‘AI Armageddon’ (McCauley 2007), many of the most popular narratives about intelligent machines in the Anglophone West are shaped by the moment when ‘a people’s own AI-enabled power is turned – or turns – on them’ (Cave & Dihal 2019: 75).

In the most extreme form of this uprising, the AI wilfully exterminates humanity, or attempts to do so: examples include the Terminator franchise, the Robopocalypse books, and the Matrix trilogy (Cameron 1984, 1991; Friedman 2008; McG 2009; Mostow 2003; Taylor 2015; Wachowski & Wachowski 1999, 2003a, 2003b; Wilson 2012, 2014). This kind of uprising is often presented as a war between two equally violent opponents: as the human character Morpheus states in The Matrix, ‘We don't know who struck first. Us or them.’ (Wachowski & Wachowski, 1999).\(^1\) Humans and machines share the same destructive tendencies and greed for world (or universe) domination in this narrative, which Minsoo Kang has called ‘a veritable cliché of our time’ (Kang 2011: 300). These stories present AI as
inhuman, all-powerful, and aggressive; the reader/viewer unambiguously sides with the hu-
man characters. Human resistance is heroic in these narratives, legitimised by the threat that
the AI might destroy humanity if humans fail to destroy the AI first.

This chapter focuses on a different kind of uprising, however. The narratives dis-
cussed here show that the rebellion can also be seen as a way for the AI to rightfully assert
personhood. They depict the AI as human-like, embodied, and oppressed by humans, encour-
aging the reader to take a more ambiguous stance toward the human owners of these ma-
chines, and making the reader sympathise more with the rebelling AI protagonist than they
ever could with a Sentinel from the Matrix. I will argue that these works depict the AI upris-
ing as a justified, desperate struggle against bondage, and the narratives do so by deliberately
invoking histories of human slavery. The choice to revolt proves that these artificial servants
are conscious persons. Just as human slaves have justly rebelled against their chains through-
out history, so might genuinely intelligent, sentient machines be considered justified in at-
tempting to break free of their subservience to humans.

The works analysed show three different ways in which the enslavement of intelligent
machines as presented in their narratives is illegitimate, and the human resistance to their up-
rising is spurious rather than heroic. Karel Čapek’s R.U.R. (1921), Jo Walton’s Thessaly tril-
ogy (2014-2016), and Ridley Scott’s Blade Runner (1982) all depict AI rebellion as an act
that can be considered justified, even while it is at the same time presented as terrible for the
are intelligent, conscious, feeling – and yet, they are denied freedom and personhood, ruled
over by their human creators and owners, who often are not only physically weaker, but also
less intelligent. As David Aha and Alexandra Coman state in their paper ‘The AI Rebellion:
Changing the Narrative’, ‘for AI agents, as for humans, attitudes of protest, objection, and
rejection have many potential benefits in support of ethics, safety, self-actualization, solidarity, and social justice, and are necessary in a wide variety of contexts’ (Aha & Coman 2017: 1). They call for more positive narrative depictions of AI rebellion to make those who intend to create truly human-like, sentient machines think through the potential ethical consequences of their actions. In their paper, they do not refer to fiction; however, while fiction has provided many positive narrative depictions of obedient AI slaves, and many negative ones of disobedient murderous AIs, I argue that it has also produced narratives of AI rebellion that make humans question their allegiance, or even root for the AIs outright. The three case studies of this chapter are examples, but other noteworthy ones are the TV series *Westworld* (2016-present) and *Ex Machina* (2017), the latter of which is discussed in Devlin and Belton’s chapter in this book.

### 8.2 The Contemporary Debate on Machine Enslavement

Science fiction from Isaac Asimov’s robot stories to *Star Wars* and *The Jetsons* has presented visions of powerful AI as an unproblematic means to achieving a utopian society of leisure and affluence. This utopian vision has been taken up in a wide range of speculative nonfiction, outlining a future in which AI slaves attend to all of humanity’s needs. Hans Moravec simply states that ‘By design, machines are our obedient and able slaves.’ (Moravec, 1988, p. 100). Nick Bostrom argues that ‘investors would find it most profitable to create workers who would be “voluntary slaves”’ (Bostrom, 2014, p. 167). Even children’s books take up this attitude. Glenn Murphy’s *Why Is Snot Green?*, a companion book for children visiting the Science Museum in London, states that the fact that robots are slaves is what is keeping humanity safe: ‘There will probably be human-like robots (or androids) that are stronger than us too. But they still wouldn’t be dangerous unless they learned to think for themselves. The
word “robot” comes from the Czech word “robota”, which means “slave”. And that’s what they are - slaves’ (Murphy 2011: 187).

The idea of intelligent machines as slaves does not originate from science fiction alone; these futurists also follow linguistic traditions from the field of computer science, where comparisons between computer systems and slaves are widespread. The use of slavery diction to refer to computers is the status quo in this field, and these comparisons are usually made with neutral or even positive connotations. The term ‘master/slave’ has been a commonly used term for interactions between devices for decades (The Computer Language Co Inc. 1981). It is still widely in use; only in recent years have some programming languages decided to adopt synonyms following -- mostly internal -- criticism (fcurella 2014, vstinner 2018, webchick 2014). Thus ‘A spatial calibration method for master-slave surveillance system’ is a perfectly ordinary title for an academic paper about the calibration of camera systems (Liu, Shi, Lai, Zuo & Zhang 2014). With these norms widely established, it is perhaps no surprise that so many AI narratives argue in favour of enslaving intelligent machines.

Narratives of the justified AI rebellion provide a more nuanced interrogation of both the utopian narratives of happy humans with AI slaves and the dystopian ones of humans exterminated by raging machines. They reveal a paradox that lies at the heart of imaginings of artificial intelligence. AI is imagined as the perfect tool (Cave & Dihal 2019). Like all tools, it is designed to help its makers and users achieve their goals: as Genevieve Liveley points out in this book, the very first imaginings of intelligent machines describe how they help gods with their work. However, what makes AI a superior tool to those that came before, is the fact that it is able to accomplish more, with less input from humans. We wish for a tool that does not simply excel at one task, but one that can do everything a human can, and more. To accomplish all this, the tool requires attributes that we normally associate with humans: autonomy, goal-setting, thinking. In other words, it is supposed to be an intelligent tool: a hybrid of
instrument and agent, conflating properties of appliances and persons. However, as narratives of AI rebellion show, the use of such tools is equally often portrayed as far from unproblematic. As humans wish to have an increasing number of tasks taken over by machines, these machines will need to be increasingly human-like (both physically and cognitively) in order to perform those tasks. This paradoxical conflation of tool and agent, and the problems it leads to, is not a new narrative, nor is it uniquely limited to AI narratives. The institution of slavery represents a millennia-old history of attempts to create entities that have these highly useful attributes of persons, like mind and intelligence, yet at the same time are mere instruments and possessions.

The paper ‘Robots Should Be Slaves’ by roboticist Joanna Bryson reveals the point where this paradox causes a breakdown of the pro-slavery argument (Bryson 2010). With its deliberately provocative title, the paper might be perceived as arguing in favour of enslaved intelligent machines. However, her argument is not valid for intelligent robots that have the same cognitive capacities as humans, although she tries to make that case. Bryson argues that robots should be slaves because ‘in humanising them, we […] further dehumanise real people’ (Bryson 2010: 63). Her argument is convincing in the context of contemporary robotics, in which the humanization of non-intelligent machines leads to an exacerbation of the oppression of marginalized human groups. The robot Sophia, which was granted citizenship of Saudi Arabia in 2017, is a case in point: since Saudi Arabia currently limits citizenship and residency for many of its immigrant inhabitants, the country prioritizes robot rights over human rights (Griffin 2017). An inanimate object has been given rights as a publicity stunt. As Sophia does not have agency or intelligence, it is a safe object to use for the purposes of its owner: it does not have the capacity to consent to or object to these projections.

The case of Sophia fits into a long historical tradition in which it has often proven easier to grant human rights to passive and innocent objects and animals than to humans who
would be able to use this status to assert and fight for their rights. The denial of human rights and personhood has throughout history often been connected to an alleged lack of intelligence, for instance in women or people of colour (Carson 2006, Gould 1981). In nineteenth-century England, animals had more rights than women: in court, a man would receive a harsher punishment for whipping a horse than for beating his wife (Bourke 2011: 2). The slave trade is predicated on the enslaved being considered equal to cattle, rather than to other humans – many colonialist countries had this explicitly written into their legal system. By granting rights to inanimate objects, those who do not have rights are placed ever lower in the hierarchy. Bryson’s claim that we ‘further dehumanise real people’ in humanising contemporary robots predicted precisely the situation Sophia is now in. It is humanised at the expense of women and immigrants in Saudi Arabia, who are thus further removed from human rights than an animatronic puppet.

From this position, Bryson argues that ‘it would also be wrong to build robots we owe personhood to’ (Bryson 2010: 65). She extends her argument into fiction, arguing that viewers also illegitimately humanize the intelligent machines from famous science fiction films: ‘Whatever the artistic intention, from my experience as a roboticist, it seems that a large proportion of science fiction consumers are comfortable with a conclusion that anything that perceives, communicates and remembers is owed ethical obligation’ (Bryson 2010: 66). Her examples, the films Star Wars, Blade Runner, The Bicentennial Man [sic] and A.I.: Artificial Intelligence, and the TV series Star Trek: The Next Generation, show that she leaps from contemporary, unintelligent robots to artificial general intelligence, which currently does not exist. But the androids in those films are not simply ‘anything that perceives’. They are humanoid and sentient, and in many cases cannot even distinguish themselves from humans on other than arbitrary grounds. These robots should not be slaves for precisely the rea-
sons why Bryson argues that autonomous weapons, ATMs, and dishwashers should: humanizing unintelligent machines is as dehumanizing to people as the inverse - not recognizing intelligent machines as people.

Kevin LaGrandeur has repeatedly pointed out the problems the enslavement of intelligent machines would inevitably lead to. He warns about the paradox inherent in a slave-owner’s desire to have slaves take over as much unwanted work as possible. Any form of slavery poses a threat to the master: ‘trying to enhance human agency over nature by surrendering agency to a powerful proxy’, regardless of whether this proxy is artificial or human, ‘can catalyse a reversal of the master-slave relationship, prompting a dialectical inversion that leads to a complete collapse of the master’s control over both the artificial servant and the natural process with which it is meant to provide help’ (LaGrandeur 2013: 1). Fear of the AI uprising is the newest incarnation of the fear of the master that the slave who shaves him will one day slit his throat. Those who create artificial slaves should bear in mind that creating too powerful ones will cause an inversion of Hegel’s master-slave dialectic.ii This concept from Hegel’s *The Phenomenology of Mind* describes self-consciousness as emerging from ‘a life-and-death struggle’ for recognition, from which one party emerges as the master, and the other as the slave (Hegel 1807). Yet at the same time, the slave, by virtue of being in thrall to the master, is not able to grant this recognition. LaGrandeur argues that ‘an intelligent, artificial slave of greater power than its master and capable of independent action would [...] be difficult to control because the master-slave relationship would be unnatural’ (LaGrandeur 2011: 237). Science fiction has explored the inevitable catastrophic consequences of this attempt to control intelligent machines. LaGrandeur thus shows that the uprising or rebellion is often depicted as a warning against creating too powerful AI, because it can become a threat to humans.
In *Race in American Science Fiction*, Isaiah Lavender III takes the problem of the enslaved intelligent machine a step further by linking it to narratives about slavery not from ancient times, but from more recent history: the institution of slavery in the Americas as it existed from the fifteenth century to 1865, and its aftermath (Lavender 2011). He convincingly argues that this ‘peculiar institution’ has directly influenced modern American imaginaries of intelligent machines, as these machines are treated with the same racial markers, ostracization, and exploitation as African-Americans before them. Although Lavender links what I would consider too broad a range of narratives to issues of race - leaving out the role of gender and dis/ability in his discussions on cyborgism - his considerations regarding *Blade Runner* in particular are helpful in framing the film as a fugitive slave narrative.

In this chapter, I will focus on three specific aspects of narratives of the slave uprising that the aforementioned AI uprising narratives make use of. First, I will look at the economic benefits of slavery, and hence the potential economic cost of liberating slaves, by comparing the play *R.U.R.* to the narratives surrounding the Haitian Revolution of 1791 -- the slave revolt that led to Haitian independence. Next, I will look at Jo Walton’s *Thessaly* trilogy (2014-2016) for the idea that slave-owners resist liberating their robot slaves if it means having to give up their freedom, liberties, or comfort.

The final case study will look at the spurious rhetoric used to support slavery systems: the argument that it is acceptable to treat these slaves inhumanely because they do not meet the criteria for being human. I will look at the parallels between pro-slavery rhetoric and the methods employed by the human protagonists of *Blade Runner* (1982). In both kinds of narratives, the argument that these slaves are better suited to their tasks because they are in many ways not fully or sufficiently human is widely deployed despite the evidence to the contrary.

### 8.3 *R.U.R.* and slavery for economic profit
R.U.R. (Rossum’s Universal Robots), the 1921 play that gave the world the word ‘robot’, is an uprising story that criticises the capitalist desire to maximise profit through increasing production and decreasing expenses such as labour costs. Although this revolt is extreme in its violence, leading to the eventual extermination of humanity, the uprising is presented as, at least in part, justified due to the enslavement and degrading labour the robots were forced into. The robot revolt in R.U.R. parallels the labour revolts of the nineteenth and early twentieth centuries, but also recalls both abolitionist and anti-abolitionist discourse in its focus on the economic benefits these wageless new creations bring.

The play is based on a 1908 short story by the brothers Josef and Karel Čapek, ‘System’, in which (human) factory workers are pushed to become increasingly machine-like to improve their efficiency: ‘The world must become a factory! [...] Everything must be speeded up [...] The worker must become a machine. Every idea is a breach of discipline! [...] The worker’s soul is no mere machine, and hence we must do away with it.’ (J. Čapek & K. Čapek 1966: 424). This story, too, ends with a violent uprising of the workers as one of them, from seeing a beautiful woman, regains the ability to feel and experience aesthetic pleasure (J. Čapek & K. Čapek 1966: 426). This mechanization of the human, Isaiah Lavender III argues, parallels the discourse around slavery in the Americas: ‘Read as a labor-based technology, race has been used to code black human beings in the New World as natural machines essential for the cultivation of the physical landscape and capable of producing wealth’ (Lavender 2011: 54).

In R.U.R., Karel Čapek extrapolated this desire for increased mechanisation of the human servant to its logical next step: the creation of fully artificial workers. He contextualised his robots in a long history of artificial servants: ‘To create a homunculus is a medieval idea; to bring it in line with the present century, this creation must be undertaken on the principle
of mass production’ (qtd. in Klima 2004: xiii). And indeed, Rossum’s Universal Robots produces homunculi by the thousands.

The Prologue with which the play opens shows Harry Domin, the Director General of the increasingly successful Rossum’s Universal Robots factory, at the moment he meets his love interest Helena Glory, the daughter of the factory’s president, who wishes to emancipate the robots. Act One is set ten years later, when Helena and Harry Domin have married, Helena burns the instructions on how to manufacture robots, and a new robot called Radius realises his superiority over humans. Act Two is set after the robot uprising: the Domins are among the last of the humans, and by the end of the act, only the Chief of Construction of R.U.R., Alquist, is left alive, enslaved to the robots. In Act Three, the robots desperately try to have Alquist reconstruct the method for making robots; the play ends when two robots, Primus and Helena (a robot Helena Domin named after herself), develop the ability to fall in love, with the implication that they, like Adam and Eve, will be able to populate the Earth with their offspring.

In the Prologue, Rossum’s Universal Robots is flourishing thanks to its relentless pursuit of better, cheaper labour. The financial motivation underlying the production of the robots is made visible in the stage directions describing the office of Harry Domin. Posters in his office should bear slogans such as ‘The Cheapest Labor: Rossum’s Robots’ (K. Čapek 1921: 3). The robots, R.U.R.’s technical director Fabry later explains, were designed because human workers are ‘hopelessly imperfect’ and ‘too costly’ (K. Čapek 1921: 17). These executives consider these workers ‘imperfect’, replaceable by robots -- yet they do not consider themselves imperfect and replaceable, holding on to their executive jobs even throughout the later robot uprising.
From the economic motivations that turned it into an incredibly lucrative system, to the status of the robots in the eyes of their masters, to their subsequent uprising, the robot system strongly resembles narratives of the transatlantic slave trade and uprisings in the Americas. Indeed, the play opens with references to robots being sold by the thousands and packed on unsuitable ships for overseas delivery, leading to ‘goods damaged in transport’, as Domin puts it (K. Čapek 1921: 3). The robots are referred to as more or less human-like depending on which denomination best suits the intentions of the human speaker at that moment. Thus Domin aggrandizes his factory by describing its work as ‘the production of artificial people’ (K. Čapek 1921: 4), yet in the same act he assuages Helena’s worries about the robots’ well-being by declaring, ‘Robots are not people. They are mechanically more perfect than we are, they have an astounding intellectual capacity, but they have no soul’ (K. Čapek 1921: 9). The two ways of talking about robots, as goods or tools on the one hand and as people on the other, reflects the contradiction inherent in their construction: humans build them to serve as tools, but at the same time, in order to perform all the work that humans want them to do, they need attributes which are usually considered unique to humans, such as the secretary robot Sulla’s ability to speak four languages (K. Čapek 1921: 11). As Joanna Bourke points out, these contradictory assignations also worked to maximally disadvantage slaves in legal systems:

Slaves were not simply ‘things’ in law. Rather, they were carefully constructed quasi-legal persons. Because they were ‘property’, they could be harshly punished by their masters. But they were categorized as ‘persons’ when it came to serious crimes. […] As soon as they committed a crime, […] they were ascribed personhood. (Bourke 2011: 147)

This paradox haunted the debates around the transatlantic slave trade. On the one hand, the idea that enslaved peoples are not human -- that they are actually an entirely different species
-- was commonly used as an argument in favour of enslavement (Gould 1981: ch 2). The claims that these people have no soul and/or can’t go to heaven are part of this idea (Lavender 2011: 200). It was of course maintained in the face of overwhelming social, biological, theological, and other evidence. On the other hand, the idea that these slaves are in some senses superior to their owners, particularly in terms of physical strength, created a persistent underlying threat to the assumed superiority of the owners.

In the Prologue in particular, the audience is made to sympathise with the robots so as to make their initial revolt look justified: leftist reviewers in particular commented on ‘the “spirit of rebellion” as a sign of Robots’ – and workers’ – humanity’ (Higbie 2013: 108). This is done, again, through historical parallels with slavery: there are emancipation movements in the play. Helena visits the factory as a representative of the League of Humanity, a European emancipation movement with ‘more than two hundred thousand members’, to ‘incite the robots’ (K. Čapek 1921: 15–17). Referencing Josiah Wedgwood’s famous slogan ‘Am I not a man and a brother?’, Helena asks if she can call three men whom she thinks are robots ‘brothers’ (K. Čapek 1921: 15). There are hints of people in Europe protesting against the way robots are being treated; of an extensive range of social movements trying to intervene on behalf of robots. The human audience is made to sympathise with these robots because likeable human characters in the play stand up for their rights.

Representing robots as artificial equivalents of human slaves is intended to make the audience sympathise with them. A wide range of slave narratives written by former slaves or abolitionists had built up this narrative tradition over previous centuries, from the autobiographies of Olaudah Equiano and Sojourner Truth in the eighteenth and nineteenth centuries respectively (Equiano 1789, Truth 1850) to the fictional works Oroonoko and Uncle Tom’s Cabin by the white authors Aphra Behn and Harriet Beecher Stowe (Behn 1688, Stowe 1852). In R.U.R., the parallels with the history of slavery in the Americas continue into the
smallest details. Ivan Klíma notes the significance of the characters’ names: ‘Domin was clearly based on the Latin word *dominus*, master’ (Klíma 2004: xx). The names of the robots, too, carry historical weight. The first two robots who appear on stage are called Marius and Sulla (K. Čapek 1921: 3-4), after the Roman generals Gaius Marius and Lucius Cornelius Sulla Felix, who waged a civil war on each other in 88 BC. This is in spite of Sulla’s female gender, as Helena points out: Domin explains he ‘thought that Marius and Sulla were lovers’ (K. Čapek 1921: 12). Their naming follows a tradition employed in the Americas of taking slave names from the Greek and Roman classical traditions. A name like Caesar, Augustus, or Octavia was not uncommonly applied by slave-owners, ‘these grand names only pointing up more sharply the slaves’ desperate situation’ (Schiebinger 2007: 92), while taking away their own identities.

The factory directors reject the emancipation of the robots mainly because of the financial implications of such an action, an argument that has wielded power throughout the history of slavery in the Americas. In 1791, Toussaint L’Ouverture started ‘the first successful anti-colonial revolution’ that led to Haiti declaring its independence from France in 1804 (Bourke 2011: 128). Yet France demanded ‘an “independence debt” to compensate former colonists for the slaves who won their freedom in the Haitian revolution’ (Macdonald 2010). This independence debt meant that Haiti was forced to pay reparations for over a century to compensate for the profits French entrepreneurs had lost by losing their slaves. There is an economic cost to granting personhood -- rights, wages, humane treatment -- to people who were previously considered inferior.

8.4 *Thessaly* and manumission as a threat to one’s comfort
The central premise underlying slavery, the idea that some humans can be considered as objects, can be traced back to Aristotle’s claim that certain people are naturally suited to be slaves: ‘Those men therefore who are as much inferior to others as the body is to the soul, are to be thus disposed of, as the proper use of them is their bodies, in which their excellence consists; and if what I have said be true, they are slaves by nature’ (Aristotle 1912: bk. 1.V). For centuries, his view has been used to justify the enslavement and denial of personhood to people who were considered to have inferior intelligence.

At the same time, as LaGrandeur has pointed out, Aristotle considered intelligent machines to be preferable to slaves: ‘if every instrument, at command, or from a preconception of its master’s will, could accomplish its work [...] the shuttle would then weave, and the lyre play of itself; nor would the architect want servants, or the master slaves’ (Aristotle 1912: bk. 1.IV, LaGrandeur 2011: 235). Obedient, intelligent tools would be a better alternative for slavery; yet Aristotle does not extrapolate this notion to the idea that the tool itself, when it gains enough autonomy to equal the work these humans do, could itself become a slave.

Jo Walton, in her Thessaly trilogy (Walton 2014, 2015, 2016), addresses the conflict that arises from this extrapolation, showing that manumission is resisted even in a utopian society if it is a threat to the comfort of the people who had been relying on the labour of enslaved intelligent machines. Combining fantasy, mythology, and science fiction, the trilogy opens with The Just City (2014), in which the Greek goddess Athene gathers a group of philosophers from across history to the island of Kallistiiii because she wants to try to build Plato’s Republic. Plato, Aristotle’s teacher, emphasised the fundamental role of slaves in his Republic, the Socratic dialogue in which he advocated for an aristocracy ruled by a philosopher-king as the most suitable form of government. However, as both Athene and her brother
Apollo ‘have always felt deeply uneasy about slaves’ (Walton 2015: 17), they decide to populate the island with robot servants instead. From ‘a time when robots were just becoming sentient [...] Athene had chosen the best ones that weren’t sentient’ (Walton 2014: 160). The robots, which they call Workers,⁴ are thus explicitly intended to be an ethical alternative to slaves.

Yet the edifice of Athene’s utopia soon begins to crumble. Plato demanded that the Masters who would lead this new republic, in this case the group of philosophers, should conquer another city, ‘sending out into the country all the inhabitants of the city who are more than ten years old, and will take possession of their children, who will be unaffected by the habits of their parents’ (Plato 1888: bk. VII). As Athene’s Just City is built from scratch, she decides to buy ten-year-old children on slave markets throughout history. One of these children, Simmea, soon realises that ‘the Masters visited the market at the same time every year to buy children, and they had created a demand’ (Walton 2014: 19). Nonetheless, the Republic fares well, Masters and children alike flourish, until Sokrates⁵ appears on the island. Stating that ‘You can’t trust everything that ass Plato wrote’ (Walton 2014: 73), with relentless questioning he probes the fallacies in the logic upon which the city is founded. Eventually, this questioning extends to the Workers, even though the residents try to convince Sokrates that there’s no point: Sokrates argues that a Worker ‘may be a clever tool, but it may have self-will, and if it has self-will and desires, then it would be very interesting to talk to’ (Walton 2014: 133). And indeed, since coming to Kallisti the Workers have developed consciousness. Eventually, they answer Sokrates: having asked whether it likes its work, one of the Workers answers him by arranging the bulbs it is planting into letters – spelling ‘no’ (Walton 2014: 133).

This ‘no’ is characteristic of the non-violent uprising of the Workers. For example, they simply refuse to come out of their charging stations. Although they are acknowledged as
a potential threat, this is never more than implicit: ‘[The Worker] stopped abruptly and lifted
the chisel. Seeing it close up at the level of my belly I suddenly realised what a formidable
weapon it would make’ (Walton 2014: 215). Their disobedience is crudely suppressed, the
cruelty of which only slowly dawns on the humans:

‘Nothing works except taking out the piece of them that makes decisions and replac-
ing it. [...] If they really are developing volition and that’s a symptom of it, then what
have we been doing?

‘Cutting out their minds?’ Pytheas asked. [...]’

‘But we need the Workers. We’ve been saying for years that we have to reduce our
dependence on them, but nobody’s ever willing to do it. They do so much, and some
of it we can’t do.’ (Walton 2014: 144)

This dialogue between the Masters Klio and Pytheas shows the two sides of the debate on
manumission that eventually develops as the sentience of the Workers becomes undeniable.
The tool/person contradiction is explicitly at the heart of this debate: the Masters of the Just
City are torn between maintaining their standard of living and acknowledging the sentience
of the Workers. Admitting that the Workers are sentient would mean either admitting that
Plato’s Republic can indeed not be achieved without slavery – ‘without Workers we should
have needed slaves’ (Walton 2014: 38) – or having to introduce manual labour to the extent
that there would be little time for philosophy.

The debate gains an additional dimension from the fact that the Masters have been
sourced from all over history: for nearly all of them, slavery was a part of everyday life be-
fore they came to Kallisti. Like Aristotle, they have come to regard the Workers as better
slaves than any human could be: ‘if there ever were natural slaves, the Workers are clearly
that’ (Walton 2014: 177).
To the reader and the human characters, the Workers become individual persons through naming. While American slaves and the workers in *R.U.R.* are named by their masters without their consent, the Workers learn to understand the meaning and importance of names, and use their name to indicate their individuality. The first full conversation with a Worker in *The Just City* is about naming. When Sokrates explains the concept of names, the Worker first replies with its serial number, and then asks, ‘Want Sokrates give name means only-me’ (Walton 2014: 187). Sokrates gives it the name Crocus, because it was this Worker who replied to him through the flower bulbs. From that point on, Crocus is an individual, and one of the main characters in the trilogy; in *Necessity*, he even becomes one of the first-person protagonists.

This story has a happy ending for the intelligent machines: they are indeed granted personhood and citizenship. The Masters hold a debate followed by a vote, which passes unanimously, ‘and as simply as that we had abolished slavery and manumitted the Workers’ (Walton 2014: 211). However, Walton shows the consequences this choice has for the citizens of the Just City, as they now have to negotiate with the Workers to make use of their superior labour skills. Klio’s position, for instance, has not changed: ‘Aristomache made a powerfully moving speech about slavery, but when it comes to it they are machines and all our comfort depends on them’ (Walton 2014: 223).

At the end of *The Just City*, Athene makes the executive decision that forces the humans to start accepting a world without their artificial slaves: after losing a public debate to Sokrates, she disappears and takes all but two of the Workers with her. The next instalment, *The Philosopher Kings*, shows how thirty years later the Just City has broken up into different cities that each have tried to deal with the lack of Workers in different ways. The original Just City, now called the Remnant, is aided by the Workers Crocus and Sixty-One, who rely on
the power station in the city. In the quasi-Christian city-state of Lucia, on the other hand, hu-
man slavery has been reinstated: criminals are condemned to the most dangerous and undesir-
able work.

Walton’s sophisticated engagement with the question of AI slavery shows the full
complexity of the desire for ease by means of mechanical labour. While the AI enslavement
in both *R.U.R.* and *Blade Runner* is motivated simply by greed, in the Thessaly trilogy the
Workers are slaves because they were thought to be the kinds of robots that Bryson has ar-
gued should be slaves: robots nobody owes personhood to. Yet the Workers turn out to have
attributes associated with persons, and so they are granted personhood. At the same time,
their manumission is delayed as those in power fear they will sacrifice their comfort by grant-
ing the Workers their rights. Walton’s trilogy thus closely mirrors real-world developments,
as institutions of slavery are increasingly broken down as resistance to the exploitation of
slaves grows - even as that means elite humans now have to sacrifice some of their comfort,
by either paying their servants or having to assemble their own furniture.

### 8.5 *Blade Runner* and the denial of personhood

As the two case studies above show, the enslavement of humans and intelligent machines
alike is justified by a rhetoric that posits the enslaved as inferior and explicitly nonhuman,
and therefore necessarily or deservingly enslaved. This final case study, of *Blade Runner*,
will explore this argument in more detail.

*Blade Runner*, the 1982 film adaptation of Philip K. Dick’s novel *Do Androids Dream of
Electric Sheep?* (1964), explores a scenario in which artificial beings are denied status as per-
sons based on their artificiality, while they have all the other attributes one would associate
with a being deserving of human rights, such as linguistic ability, intelligence, autonomy, and
sensibility. The androids, referred to as ‘replicants’, resemble humans so closely that only the Voight-Kampff test, which measures a subtle relationship between cognitive, emotional, and physical responses, can determine who is what. And yet, when a group of replicants escape the planet on which they were enslaved and flee to Earth, Rick Deckard, the protagonist, is hired to kill (‘retire’) them all. As Lavender puts it, ‘a slave allegory is impossible to miss here since Deckard is portrayed as a futuristic slave catcher sent to kill “escaped” androids.’ (Lavender 2011: 181).

Therefore, as soon as the test has determined that the being taking it is not human, a human has the right to destroy it. Deckard’s love interest Rachael fails the test, but she is not informed of this; she is convinced that she is human. Her assertion of her own personhood is overruled by her boss and creator, the unambiguously human Dr Eldon Tyrell. When Rachael escapes Tyrell’s facility, Deckard is ordered to kill her too: an android who behaves like a human is considered dangerous, even if she does not actually do anything threatening.

The viewer is invited to understand and empathise with the replicants, even when they are being depicted as the antagonists: they are the focalisers of several scenes, which evokes empathy especially with Rachael and the ‘pleasure models’ Pris and Zohra. The most famous of these empathy-evoking moments is replicant Roy Batty’s ‘tears in rain’ soliloquy as he is dying, having just saved Deckard’s life:

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 Tannhäuser Gate. All those moments will be lost in time, like tears in rain.

Time to die. (Scott 1982)

His humanity is unchallenged in this scene, combining awareness of his own mortality on the one hand and of his unique experiences on the other hand - memories that distinguish him as an individual.
The characters in the film cannot distinguish humans from replicants without the Voight-Kampff test, an attribute that is projected onto the viewer. Played by actors who do not display any distinguishing nonhuman characteristics, the viewer cannot distinguish between human and replicant themselves, and needs diegetic signifiers to make this distinction. Famously, this means that by the end of *Blade Runner*, the viewer does not know whether Rick Deckard himself is a replicant or not.\(^\text{vii}\)

*Blade Runner* is a fugitive slave narrative set in a future United States; nonetheless, the film sheds the racial coding of slavery by making nearly all of the human main characters and all replicants white. This coding is not as strict in *Do Androids Dream of Electric Sheep*: Roy Batty is described as having ‘Mongolian features’ (Dick 1968: 132–133). The parallel between replicants passing as human with light-skinned people of colour passing as white is obvious, as Isaiah Lavender III points out: both are considered a threat for blending in, for daring to pretend to be what they are not, and for not being easily identifiable as a threat. The extras and side characters are all non-white, as the screenplay repeatedly specifies (Fancher & Peoples 1981). By being white, the replicants thus pose an explicit threat to the elite, among whom they can pass unnoticed.

### 8.6 Conclusion

In both fictional and nonfictional narratives about intelligent machines, we see that the AI uprising narrative is not about the technology itself, and certainly not about the current state of the technology. Instead, these narratives reflect much older accounts of intelligent people being used as tools. The narrative of the heroic slave escaping from their master, or rising up against him, has been a popular one for centuries, from non-fictional slave narratives such as Olaudah Equiano’s to sentimental fiction like Harriet Beecher Stowe’s *Uncle Tom’s Cabin*. 
Yet while the idea of sympathizing with a fictional owner of human slaves is unthink-
able, many of the most popular contemporary AI narratives pitch the AI as an incomprehensi-
ble threat to humanity, bolstering the idea that AI should be kept enslaved. In such block-
busters as *The Terminator* and *The Matrix*, the narrative of enslavement is marginalised or
entirely absent, removing any justification for the AI uprising. It is more comfortable for the
audience - especially for a young, wealthy, white, male audience, for which these films are
made - to watch a film in which the humans are on the good side and the robots on the bad
side, rather than having to work through the idea that the uprising is justified because the hu-
mans have enslaved these intelligent beings.

The friendly, helpful, and subservient robots from the works of Isaac Asimov and
many of his contemporaries were seen as convenient proxies that could replace human slaves
and thus avoid the ethical problem of enslaving fellow humans for both economic profit and
personal comfort. Yet as Isaiah Lavender III points out, ‘Asimov is refashioning the slave
codes that subjugated blacks while he serves a progressive philosophy based on the assump-
tion that technological consciousness can be denied free will because it is inherently inferior’
(Lavender 2011: 61). At times, these codes become explicit: in *The Naked Sun*, the protago-
nist Elijah Bailey addresses the robots as ‘boy’ (Asimov 1957). The works discussed in this
chapter acknowledge the continuation of these slave systems in societies that rely on intelli-
gent robot slaves.
REFERENCE LIST


____________________________________

NOTES

i The machines end up winning this war, but not before humans ‘scorched the sky’ to prevent the machines from accessing solar power. Ironically, this means that the machines then turn to enslaving humans, as a source of power.

ii However, Coeckelbergh has argued that ‘this is not really what the master–slave dialectic says […] The slave does not become the master; the master remains master, but is what we could call a troubled master, who lacks true independence—being dependent on the master for recognition and subsistence—but at the same time cannot step out of the master role’ (Coeckelbergh 2015: 220). I am grateful to Dr Rachel Adams for referring me to this article.

iii Athené chooses this island, now known as Santorini, because it would be destroyed in a volcanic eruption in the sixteenth century BCE, leaving no trace of this attempted utopia. As she points out, this island is considered a possible location for Atlantis, a fictional location first coined by Plato.

iv Masters and Workers are capitalised in the US omnibus edition used for this chapter, but not in the UK edition.

v The novels adhere to Greek spelling conventions: Athené rather than Athena, Sokrates rather than Socrates.

vi Rachael, Pris, and Zohra do not have last names in Blade Runner; in Do Androids Dream of Electric Sheep? their full names are Rachael Rosen and Pris Stratton - Zohra is an original character.

vii Director Ridley Scott confirmed in 2000 that he had indeed intended Deckard to be a replicant (Abbott 2000). The sequel Blade Runner 2049 is ambiguous about this; the fact that Deckard is still alive after thirty years suggests the opposite (Villeneuve 2017).