

# S AVANNAH LAW REVIEW

VOLUME 5 | NUMBER 1

## “I AM THE MASTER”<sup>v</sup>: SOME POPULAR CULTURE IMAGES OF AI IN HUMANITY’S COURTROOM

---

*Christine A. Corcos\**

### Introduction

Both serious literature and popular culture are flooding us with discussions of the rise of artificial intelligence (AI).<sup>1</sup> As we note the rise of the subject of robot

---

<sup>v</sup> Gnut, in Harry Bates, *Farewell to the Master*, ASTOUNDING SCIENCE FICTION (Oct. 1940), reprinted in ISAAC ASIMOV PRESENTS THE GOLDEN YEARS OF SCIENCE FICTION (1988).

<sup>\*</sup> Richard C. Cadwallader Associate Professor of Law, Louisiana State University Law Center, Associate Professor of Law, Louisiana State University A&M. I presented the paper on which I based this article at *Savannah Law Review*'s Annual Colloquium, *Rise of the Automaton*s, September 15, 2017, and I thank the other panelists, the faculty, staff, students, and other attendees at the Colloquium for their helpful comments.

<sup>1</sup> Some recent important non-fiction works published on the subject include LUKE DORMEHL, *THINKING MACHINES: THE QUEST FOR ARTIFICIAL INTELLIGENCE* (2017), AMIR HUSAIN, *THE SENTIENT MACHINE: THE COMING AGE OF ARTIFICIAL INTELLIGENCE* (2017), and MAX TEGMARK, *LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE* (2017). Popular articles intended for the non-specialist include Gary Shteyngart, *Thinking Outside the Bots*, SMITHSONIAN, June 2017, at 66. We can pinpoint the first appearance of the word “robot” in popular culture to Karl Capek’s play *R.U.R.* (Rossum’s Universal Robots, Prague: Aventinum, 1920) (Claudia Novack-Jones trans., reprint Penguin 2004). The word “robot” derives from Old Slavonic (and Czech words) “robota” meaning “forced labor” or “work.” See *Science Diction: The Origin of the Word “Robot”*, NPR (Apr. 22, 2011), <http://www.npr.org/2011/04/22/135634400/science-diction-the-origin-of-the-word-robot>; see also James D. Naughton, *Futurology and Robots: Karel Capek’s R.U.R.*, 28 RENAISSANCE AND MOD. STUDIES 72 (1984). For an attractively illustrated history of robots, see BEN RUSSELL, *ROBOTS: THE 500-YEAR QUEST TO MAKE MACHINES HUMAN* (2017).

law<sup>2</sup> and particularly the question of whether AI could possibly become sentient we begin to take seriously concerns about the regulation of the use of robots and the possibility that AI might pose a threat to the physical safety and privacy of human beings. In particular, we are beginning to wonder how we might control this new technology, which seems both more intelligent and more powerful than human beings. Suppose unethical or negligent programmers create situations in which AI escapes human controls and thus contravenes human norms or rules? Can we bring that AI to account? Ought we to do so, particularly if that AI is sentient or approaches sentience?<sup>3</sup> At first, we might think that the answer should be “yes,” because after all we have created the AI and we should continue to control it.

But the question is, I would submit, more complicated. We have created computers and robots as useful tools, but we have continued to develop them as far more—as devices that far outstrip our own capacities to decipher the mysteries of the Universe. If we deliberately endow them with characteristics that mimic our own, if they develop those independently, or develop others by analogy allowing them to function in ways that mirror human activities, can we continue to insist that we should treat them as property and that they should do our bidding? If at some point they make some demand for the right not to follow commands that we issue, for whatever reason, ought we to ignore that demand?

Novelists, filmmakers, and other artists who create popular culture have already considered this question for decades, if not centuries. In this Article, I discuss some of the ways in which some of them have thought about these issues and the insights they have had, which could guide us as we move through this important area.

## I. Defining AI

### A. Real World Definitions of AI

In the real world, AI is currently not nearly as highly developed as film, television, and novels portray it. Deep Blue,<sup>4</sup> its successor Watson,<sup>5</sup> and other real-world AI demonstrate their abilities to dominate humans in the areas of game

---

<sup>2</sup> A sampling of the rapidly growing bibliography of treatises on robot law includes: RYAN CALO & A. MICHAEL FROOMKIN, *ROBOT LAW* (2016); GABRIEL HALLEVY, *WHEN ROBOTS KILL: ARTIFICIAL INTELLIGENCE UNDER CRIMINAL LAW* (2013); UGO PAGALLO, *THE LAWS OF ROBOTS: CRIMES, CONTRACTS, AND TORTS* (2013).

<sup>3</sup> I discuss the meaning of “sentience” in more detail in part II of this Article, but, in brief, by “sentience” I mean that the being is self-aware. It is “conscious of [its] own character, thoughts, emotions, . . . aware of the fact of its own existence.” *Self-Aware*, OXFORD ENGLISH DICTIONARY, <http://www.oed.com/view/Entry/57122363?redirectedFrom=self+aware#eid> (last visited September 4, 2018).

<sup>4</sup> Steven Levy, *What Deep Blue Tells Us About AI in 2017*, WIRE (May 23, 2017), <https://www.wired.com/2017/05/what-deep-blue-tells-us-about-ai-in-2017/>.

<sup>5</sup> Will Grunewald, *FYI: Which Computer Is Smarter, Watson or Deep Blue?*, POPULAR SCI. (Dec. 12, 2012), <https://www.popsci.com/science/article/2012-12/fyi-which-computer-smarter-watson-or-deep-blue>.

playing. Alexa and other robots move into homes, offering automated assistance for harried humans.<sup>6</sup> The Internet of Things (IoT) becomes part of everyday life, leading to serious discussion of how to regulate AI.<sup>7</sup> What happens when those robots that clean our homes also collect data about our lives?<sup>8</sup> What if the self-driving cars that take us to work every day make errors that cause accidents?<sup>9</sup> Although I do not propose in this Article to discuss what AI is or whether or how to regulate it, given the extent to which such issues dominate academic, governmental, and popular discussion, we must confront these questions quite soon.<sup>10</sup> At least one government has already granted one robot a right that many humans are desperate to obtain. Saudi Arabian officials announced that their government had granted citizenship to Sophia, the AI robot from Hanson Robotics,<sup>11</sup> at an event held in Riyadh in October, 2017.<sup>12</sup>

---

<sup>6</sup> Will Oremus, *Terrifying Convenient*, SLATE (Apr. 3, 2016), [http://www.slate.com/articles/technology/cover\\_story/2016/04/alexa\\_cortana\\_and\\_siri\\_aren\\_t\\_novelties\\_anymore\\_they\\_re\\_our\\_terrifyingly.html](http://www.slate.com/articles/technology/cover_story/2016/04/alexa_cortana_and_siri_aren_t_novelties_anymore_they_re_our_terrifyingly.html).

<sup>7</sup> Jeddiah Bracy, *Senate Committee Explores Internet-of-Things Regulation*, THE PRIVACY ADVISOR (Feb. 12, 2015), <https://iapp.org/news/a/senate-committee-explores-internet-of-things-regulation/#>; Anne Hobson, *Regulating "Internet of things" Requires a Better Definition From Lawmakers*, THE HILL (Jan. 24, 2017), <http://thehill.com/blogs/pundits-blog/technology/315915-regulating-internet-of-things-requires-a-better-definition-from>; Sam Thielman, *Acting Federal Trade Commission Head: Internet of Things Should Self-Regulate*, THE GUARDIAN (Mar. 14, 2017), <https://www.theguardian.com/technology/2017/mar/14/federal-trade-commission-internet-things-regulation>; Adam Thierer & Andrea O'Sullivan, *Leave the Internet of Things Alone*, U.S. NEWS (June 12, 2017), <https://www.usnews.com/opinion/economic-intelligence/articles/2017-06-12/dont-stifle-the-internet-of-things-with-regulation>; Rob Wright, *Bruce Schneier: It's Time For Internet-of-Things Regulation*, THE HILL (Feb. 15, 2017), <http://thehill.com/blogs/pundits-blog/technology/315915-regulating-internet-of-things-requires-a-better-definition-from>.

<sup>8</sup> Maggie Astor, *Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Shared*, N.Y. TIMES (July 25, 2017), [https://www.nytimes.com/2017/07/25/technology/roomba-irobot-data-privacy.html?\\_r=](https://www.nytimes.com/2017/07/25/technology/roomba-irobot-data-privacy.html?_r=).

<sup>9</sup> Neal E. Boudette, *Tesla's Self-Driving System Cleared in Deadly Crash*, N.Y. TIMES (Jan. 19, 2017), <https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html>; Michael Hiltzik, *Self-Driving Car Deaths Raise the Question: Is Society Ready For Us To Take Our Hands Off the Wheel?*, L.A. TIMES (Apr. 3, 2018), <http://www.latimes.com/business/hiltzik/la-fi-hiltzik-self-drive-20180403-story.html>.

<sup>10</sup> See Christopher Fonzone and Kate Heinzelman, *Should the Government Regulate Artificial Intelligence? It Already Is*, THE HILL (Mar. 26, 2018), <http://thehill.com/opinion/technology/375606-should-the-government-regulate-artificial-intelligence-it-already-is>; Carlos Melendez, *Should AI Be Regulated*, INFO WORLD (Apr. 3, 2018), <https://www.infoworld.com/article/3267609/artificial-intelligence/should-ai-be-regulated.html>; Carlos Perez, *Why Elon Musk is Right About AI Regulation*, MEDIUM (July 29, 2017), <https://medium.com/intuitionmachine/why-elon-musk-is-right-about-ai-regulation-7638192b4cdb>.

<sup>11</sup> Sophia, HANSON ROBOTICS (2017) <http://sophiabot.com/about-me/> (last visited September 4, 2018).

<sup>12</sup> Andrew Griffin, *Saudi Arabia Grants Citizenship to a Robot For the First Time Ever*, THE INDEPENDENT (Oct. 28, 2017), <https://www.independent.co.uk/life-style/gadgets-and-tech/news/saudi-arabia-robot-sophia-citizenship-android-riyadh-citizen-passport->

Part of our difficulty arises because we still do not have a mature, consistent, and fully agreed upon definition of what constitutes “AI.” The 2016 National Science and Technology Council Report on Artificial Intelligence, generally referred to as the “Obama Report,” notes that a number of taxonomies exist to categorize AI. One useful one divides AI into descriptive categories “(1) systems that think like humans . . . (2) systems that act like humans . . . (3) systems that think rationally . . . and (4) systems that act rationally . . . .”<sup>13</sup> Depending upon which system of AI we are discussing and the uses we make of it, we may have different expectations and different rules for that AI. We might not even agree that the system *is* AI.

### B. How AI Is Presented in Popular Culture

Popular culture presents AI in numerous ways but generally as either equal to or overwhelmingly superior to human intelligence, and thus ultimately and eventually as a threat to human survival. In one particularly stark graphic representation, AI is overwhelmingly dominant over and malevolent toward humanity, showing how AI in our daily lives could betray us, how AI in the workplace can destroy our normal expectations, and how AI, which should defend us, can instead take over and dominate us.<sup>14</sup> As in literature and films that feature other non-human characters, such as vampires,<sup>15</sup> aliens,<sup>16</sup> animals,<sup>17</sup> and

---

future-a8021601.html. One can dismiss the grant as a public relations move, but it is still extremely interesting, and a grant of citizenship, given what it represents in terms of the relationship between the individual and the state, is quite significant.

<sup>13</sup> EXECUTIVE OFFICE OF THE PRESIDENT, NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, COMMITTEE ON TECHNOLOGY, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 6-7 (2016), [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf) (citing STUART RUSSELL & PETER NORVIG, *ARTIFICIAL INTELLIGENCE: A MODERN APPROACH* (3d ed. 2009)).

<sup>14</sup> Jeff Desjardins, *The Future of Artificial Intelligence, According to Pop Culture*, VISUAL CAPITALIST (Sept. 16, 2017), <http://www.visualcapitalist.com/future-ai-pop-culture/>. Infographic first published at *How Worried Should You Be About Artificial Intelligence?*, BBC (July 14, 2015), <http://www.bbc.com/future/story/20150713-how-worried-should-you-be-about-artificial-intelligence>.

<sup>15</sup> BRAM STOKER’S DRACULA (Columbia Pictures 1992) (vampire-themed stories and TV shows that argue that vampires and werewolves have rights (e.g., *True Blood* (HBO television broadcast 2008–2014) and *Forever Knight* (Glen Warren Productions television broadcast 1992–1996) (800-year-old vampire works as Toronto police officer and tries not to exploit the humans around him)).

<sup>16</sup> ALIEN NATION (Twentieth Century Fox Television 1988); *Alien Nation* (FOX television broadcast, 1989–1990) (alien ship crashes in Mojave Desert; U.S. grants survivors refugee status); see Christine A. Corcos, *Visits to a Small Planet: Rights Talk in Some Science Fiction Film and Television Series From the 1950s to the 1990s*, 39 STETSON L. REV. 183, 209–29 (2009).

<sup>17</sup> JAWS (Universal Pictures 2005) (peaceful Maine vacation town takes up arms against a “rogue shark”).

monsters,<sup>18</sup> humans often take the position that AI that threatens their dominance in the universe is a direct threat to human existence.

This kind of representation certainly began as far back as the real-life automatons of the 18<sup>th</sup> century, which their promoters and (supposed) creators presented as wonders that could rival humans in very limited spheres (for example, in chess).<sup>19</sup> Because real AI was undeveloped at the time, automatons were objects of deceit at the same time that they were creations of wonder. Yet AI was also a subject of the imagination; it fascinated artists because it offered them an opportunity to fantasize about the creation story.<sup>20</sup>

## II. Philosophical Assumptions, Legal Regimes, and Extra-Legal Norms

We generally anchor the idea that humans have the right to punish AI if it transgresses human laws in the assumption that because humans make the laws they also control the enforcement of those laws.<sup>21</sup> This assumption follows naturally from two others: humans create the AI and humans create the conditions

---

<sup>18</sup> MARY WOLLSTONECRAFT SHELLEY, *FRANKENSTEIN; OR, THE MODERN PROMETHEUS* (Gutenberg Project ed., 2008) (1818).

<sup>19</sup> One of the most famous automatons was the 18<sup>th</sup> century inventor Wolfgang von Kempelen's Mechanical Turk, which he billed as a machine that could play chess and defeat human opponents without the assistance of a human agent. The famed magician Robert-Houdin offered the explanation that someone hid inside and operated the machine's levers and thus played the game. See TOM STANDAGE, *THE TURK: THE LIFE AND TIMES OF THE FAMOUS EIGHTEENTH-CENTURY CHESS-PLAYING MACHINE* 92–94 (2003); Ella Morton, *Object of Intrigue: The Turk, a Mechanical Chess Player That Unsettled the World*, ATLAS OBSCURA (Aug. 18, 2015), <http://www.atlasobscura.com/articles/object-of-intrigue-the-turk>.

<sup>20</sup> SHELLEY, *supra* note 18. An example of a creation story relevant to this discussion is *Frankenstein*, which has developed its own bibliography. Mary Shelley was, however, also concerned with what makes us human, as was Philip K. Dick, the author of *Do Androids Dream of Electric Sheep?*, the basis for the film *Blade Runner*. See Aaron Barlow, *Philip K. Dick's Androids: Victimized Victimizers*, in *RETROFITTING BLADE RUNNER: ISSUES IN RIDLEY SCOTT'S BLADE RUNNER AND PHILIP K. DICK'S DO ANDROIDS DREAM OF ELECTRIC SHEEP?* 77 (Judith B. Kerman ed., 1991).

<sup>21</sup> We call the argument that humans should have primacy over other animals and thus should make all decisions regarding the organization, government, and use of Earth “human exceptionalism” or “anthropocentrism.” In his essay *Minding the Animals*, Steven Best explains the origin of this idea and attacks the notion that we should accept the human “model of intelligence and communication” as the norm by which we judge all others.

The argument of cognitive ethology is not that animal emotions and consciousness are as complex as ours, but that they exist in remarkably rich forms. Human beings are unique in the degree to which they possess intelligence; no other species, to my knowledge, has written sonnets or sonatas, solved algebraic equations, or meditated on the structure of the universe. But humans are not unique in their possession of a neocortex; of complex emotions like love, loneliness, empathy, and shame; of sophisticated languages, behaviors, and communities; and perhaps even of aesthetic and moral sensibilities.

Steven Best, *Minding the Animals*, <http://www.drstevebest.org/MindingTheAnimals.htm> (last visited Apr. 24, 2018).

under which the AI functions. Human developers are currently debating the origins of AI moral agency, including whether humans can assign such ethics or morals to AI or whether AI can develop them independently.<sup>22</sup> In addition, humans are attempting to determine whether AI can or should develop the conditions under which AI should perform its activities. If AI develops its own moral or ethical norms, even if those norms resemble human norms, that development will challenge human society in many ways.

That humans would relinquish control of law-making or allow non-humans, albeit represented by humans, to participate in law-making, is generally a difficult proposition for humans to accept.<sup>23</sup> Some philosophers reject completely the idea

---

<sup>22</sup> Philosophers, analysts, and, most recently, lawyers are now discussing whether humans should imbed or “teach” AI morality (which it then might continue to develop on its own), or whether AI could develop such norms independently. For a general discussion of the issues involved in teaching morality to AI, see Jane Zavaishina and Vyacheslav Polonski, *Teaching Morality to Machines*, Council on Foreign Relations, (Nov. 14, 2017), <https://www.cfr.org/blog/teaching-morality-machines>.

Germany’s Federal Ministry of Transport and Digital Infrastructure Ethics Commission has published a series of principles that it recommends guide the development of self-driving automobiles. Some of them address philosopher Philippa Foot’s well-known thought experiment the “trolley problem,” in which we must decide whether, or in what circumstances, we would kill one human being to save five (or five hundred) others. Human beings might make such a decision based on emotion (if the one individual is old, then killing her is a justifiable decision) or on moral principles, but AI right now has neither. The Ethics Commission principle is that humans should provide self-driving vehicles with “ethical rules” to make such decisions without concerning themselves with the potential victim’s age, gender, health (Principle 9). See *Automated and Connected Driving Report*, FED. MINISTRY TRANSPORT & DIGITAL INFRASTRUCTURE, ETHICS COMM’N at 11 (Ger.) (June 2017), [https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?\\_\\_blob=publicationFile](https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile). The document also preferences human life over the damage to AI-equipped vehicles (Principle 7).

In March 2016, Microsoft launched Tay, a chatbot that the company intended to use as an experiment to see how quickly and effectively such an automated device could learn from humans on social media. Tay learned a great deal, mostly about foul language, Nazi slogans, and anti-feminism, and Microsoft engineers pulled it offline within 24 hours. “The AI chatbot Tay is a machine learning project, designed for human engagement . . . . Unfortunately, within the first 24 hours of coming online, we became aware of a coordinated effort by some users to abuse Tay’s commenting skills to have Tay respond in inappropriate ways,” a Microsoft spokesperson noted. James Risley, *Microsoft’s Millennial Chatbot Tay.ai Pulled Offline After Internet Teaches Her Racism*, GEEKWIRE (Mar. 24, 2016, 9:30 AM), <https://www.geekwire.com/2016/even-robot-teens-impressionable-microsofts-tay-ai-pulled-internet-teaches-racism/>. Tay seems to have lacked a failsafe that would have protected her from picking up offensive language. On the other hand, if her creators wanted her to learn about humans, she certainly encountered some of them, and she brought home what she learned. On the trolley problem, see Judith Jarvis Thomson, *The Trolley Problem*, 94 YALE L.J. 1395 (1985).

<sup>23</sup> At a minimum, such issues implicate questions of standing. See generally Paul Schiff Berman, *Essay: An Observation and a Strange but True “Tale”: What Might the Historical Trials of Animals Tell Us About the Transformative Potential of Law in American Culture?*, 52 HASTINGS L.J. 123 (2000); Paul Schiff Berman, *Rats, Pigs, and Statues on Trial: The Creation of Cultural Narratives in the Prosecution of Animals and Inanimate Objects*, 69 N.Y.U. L. REV. 288 (1994) (trying animals for transgressing human law, or

that animals, for example, have anything like equal rights.<sup>24</sup> Such a position is understandable; it emerges naturally from the theory that only certain groups in a society should have the right to govern in a society. But this position also endlessly recapitulates the favorable position of those whose rights that society already recognizes.

The short story “Farewell to the Master,” the basis for the film *The Day the Earth Stood Still*,<sup>25</sup> tells the story of Klaatu, an alien humanoid visitor to Earth, who brings the news that other sentient species in the universe have judged humans to be dangerously off-track in their behavior, warlike and uncivil. The alien visitor, accompanied by a green metal robot named Gnut, tells the humans to alter their behavior or risk destruction. The humans react in various but entirely predictable ways, and end by killing the alien. The narrator of the story, reporter Cliff Sutherland, finally addresses Gnut to apologize, asking it to tell its masters (its creators) that Earth’s people are peaceful. Gnut listens carefully and then tells Sutherland, “You misunderstand. I am the master.” Sutherland, like the rest of the human population, has assumed that the humanoid with whom they have been interacting is the leader of the expedition, a member of the dominant species from another world. Already amazed that there are other intelligent beings in the universe, the humans have adjusted their philosophy and to some extent their views of their origins. However, they have never considered the idea that a non-humanoid might be “in charge” of another intelligent life form or its society.

Similarly, in many science fiction (“SF”) films, TV episodes, and in fiction, we encounter the same assumptions. If other intelligent life forms exist in the universe, they will be like us, or at least understand us and our values.<sup>26</sup> Humans

---

simply blaming them for disasters or for their own instinctive behavior, has been of academic interest for some time and is linked to the belief in and fear of witchcraft); E. P. EVANS, *THE CRIMINAL PROSECUTION AND CAPITAL PUNISHMENT OF ANIMALS* (1906); Walter Woodburn Hyde, *The Prosecution and Punishment of Animals and Lifeless Things in the Middle Ages and Modern Times*, 64 U. PA. L. REV. & AM. L. REG. 696 (1916) (the social and cultural historian Robert Darnton describes myriad types of revenge taken against cats, including over fear of their association with suspected witches, over the centuries); see also ROBERT DARNTON, *THE GREAT CAT MASSACRE AND OTHER EPISODES IN FRENCH CULTURAL HISTORY 75-104* (1984).

<sup>24</sup> See TIBOR R. MACHAN, *PUTTING HUMANS FIRST: WHY WE ARE NATURE’S FAVORITE* (2004).

<sup>25</sup> *THE DAY THE EARTH STOOD STILL* (Twentieth Century Fox Home Entertainment 1951); see also *THE DAY THE EARTH STOOD STILL* (Twentieth Century Fox Home Entertainment 2008) (remake and loose adaptation of the original 1951 film).

<sup>26</sup> One example of the human-centric approach is the creation and inclusion of the Voyager Golden Record, which represents a range of human knowledge and natural experiences on Earth accessible through the five human senses. Timothy Ferris discusses the method for creating the record (two records). See Timothy Ferris, *How the Voyager Golden Record Was Made*, *THE NEW YORKER* (Aug. 20, 2017), <https://www.newyorker.com/tech/elements/voyager-golden-record-40th-anniversary-timothy-ferris>.

One obvious and practical reason that many popular culture aliens have resembled humans is that human actors play aliens, and makeup artists have limited ways in which to

will be able and should be able to impose ideas of order and civility on these life forms, particularly when we must preserve human life. If we need formality to do so, in the form of some legal regime, then the formality will resemble human procedures. Humans will subject non-human transgressors, even if they are sentient, to human rules.<sup>27</sup> Nowhere is this idea more obvious than in SF that involves AI—robots or androids, for example.

When we deal with popular culture that involves AI, we also make major assumptions, one of which is that humans are and should be in charge of that AI. After all, humans create AI and they direct its development. If questions arise with regard to what kind of programming it should have, humans resolve those questions among themselves. Whatever debates may arise are debates that human beings carry on. AI has nothing to say in the matter, because AI (robots, androids, computers, and similar devices) is not a player in these matters. Humans and AI are the two parties involved but currently only one of the parties is actually making the decisions.

At the same time, when we examine representations of AI in SF popular culture, we see that pop culture creators readily engage AI as one of the parties, even if they may not consider that AI is or ought to be in charge. But the question of who ought to be in charge, who ought to be making the decisions is one we ought to consider. Thus, examining some of the fictional discussions that pop culture offers us can assist us.

---

alter human anatomy. However, with the advent of computer generated imaging (CGI), creators can produce aliens that resemble absolutely nothing on Earth; writers and filmmakers are limited only by their imaginations.

However, some scientists, as well as sf television writers, attempt to explain the alien/human similarities in other ways. See Nola Taylor Redd, “*Star Trek*” *Science: Why Vulcans (and Other Aliens) Look Like Humans*, SPACE.COM (Dec. 31, 2016), <https://www.space.com/35188-star-trek-alien-evolution.html>. The question of what aliens might actually look like is beyond the scope of this article, but is a matter of some debate in both scientific and popular culture circles. See Christine Wilcox, *The Scientific Explanation For Why Humans Are So Convinced That Aliens Look Like Octopuses*, QUARTZ (Dec. 8, 2016), <https://qz.com/857377/the-aliens-in-arrival-look-like-octopuses-because-humans-think-cephalopods-are-both-scary-and-smart/>; *Why Do Aliens In Films Almost Always Look Like Humans?*, BBC (Mar. 9, 2017), <http://www.bbc.co.uk/programmes/p04wb0qs>. Also, see the very helpful entry *Human Aliens* at <http://tvtropes.org/pmwiki/pmwiki.php/Main/HumanAliens?from=Main.HumanAlien>. Indeed, some scientists now suggest that extra-terrestrial life might actually resemble human life more than we might expect. See Mike Wehner, *Scientists Now Claim Aliens Might Be a Lot Like Humans*, BGR (Nov. 1, 2017), <http://bgr.com/2017/11/01/do-aliens-exist-research-study-they-might-look-human/>.

<sup>27</sup> Consider the objections that *Star Trek* Captain Jean-Luc Picard makes to the proceeding that he and the *Enterprise* crew must submit to in *Encounter at Farpoint*, the first episode of the series *Star Trek: The Next Generation*. Q, the extremely powerful alien being the crew encounters, accuses humanity of being inferior and savage and puts them on trial to face charges. *Star Trek: The Next Generation: Encounter at Farpoint* (NBC television broadcast Sept. 28, 1987) (Picard finally admits that humanity in particular has sometimes been uncivilized, but he notes first that human beings have progressed and second that the proceedings are unfair).

Looking back to Mary Shelley's novel *Frankenstein*, we see that novelists have considered the question of whether the creator ought to have full control of his or her creation, or whether the creation ought to have something to say about its future. Strictly speaking, the monster in *Frankenstein* is not AI,<sup>28</sup> but it is a human creation outside the usual means of reproduction, and Dr. Victor Frankenstein, the creator, tries to control his creation, but soon discovers that he cannot. The creation becomes sentient and asserts its will. It develops its desires when it requests a companion, just one example of its development toward humanity.

“You must create a female for me with whom I can live in the interchange of those sympathies necessary for my being. This you alone can do, and I demand it of you as a right which you must not refuse to concede.” . . . “I do refuse it,” I replied, “and no torture shall ever extort a consent from me. You may render me the most miserable of men, but you shall never make me base in my own eyes. Shall I create another like yourself, whose joint wickedness might desolate the world. Begone! I have answered; you may torture me, but I will never consent.” “You are in the wrong,” replied the fiend, “and instead of threatening, I am content to reason with you. I am malicious because I am miserable. Am I not shunned and hated by all mankind? You, my creator, would tear me to pieces and triumph; remember that and tell me why I should pity man more than he pities me? You would not call it murder if you precipitate me into one of the ice-rifts and destroy my frame, the work of your own hands. Shall I respect man when he condemns me?”<sup>29</sup>

Note the creature's labeling of Dr. Frankenstein's willingness to destroy it as “not murder.” Even though the creature has many of the attributes of a human being, including the desire to bond with another like it, it recognizes that Frankenstein still regards it as less than human, and therefore disposable. Because he created it, he can destroy it. Further, although the creature doesn't know it, Frankenstein thinks of it as a “fiend,” and “other,” for wanting what human beings crave and receive as a matter of course—agency, companionship, respect.

Thus, the issue of whether a creature, or AI, created as a servant for its human creators, if it develops sentience and free will, ought to have the same right of action, of agency as its agency, becomes one that the SF tradition takes seriously under discussion.

---

<sup>28</sup> On the similarity of Frankenstein's monster and today's androids, robots, and replicants, see Jay Clayton, *Frankenstein's Futurity: Replicants and Robots*, in THE CAMBRIDGE COMPANION TO MARY SHELLEY 84 (2003). See also Francine Prose, *How Frankenstein's Monster Became Human*, THE NEW REPUBLIC (June 16, 2016), <https://newrepublic.com/article/134271/frankensteins-monster-became-human>.

<sup>29</sup> SHELLEY, *supra* note 18, at § 17.

### A. Models of AI Regulation

Often the reaction of both writers and readers is that humans create machines to serve them, not so the machines can dominate their creators. The fear that machines and technology will dominate or displace human beings is an old one. Consider the reactions of the Luddites.<sup>30</sup> Relatively early on in the Industrial Revolution, we see the rise of the idea that if machines threatened humans, humans could react by curbing or destroying them. Similarly, in SF, we see the beginnings of legal regimes created to control and dominate AI. The best-known and most formal of these is Isaac Asimov's *Three Laws of Robotics*, which first appears in print in 1942.<sup>31</sup>

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

Eventually Asimov added a fourth law, which he named the Zeroth Law: *A robot may not injure humanity or, through inaction, allow humanity to come to harm.*<sup>32</sup>

While the Three Laws seem to have a philosophical and practical purpose in developing and controlling robots, Asimov obviously found them of great use in driving the plots of his stories.<sup>33</sup> A good many SF AI creators seem to follow Asimov's principles and give us AI that for some reason devolves from one or more of the laws. Indeed, these do spur plots. For example, both *Colossus: The Forbin Project* and *2001: A Space Odyssey*<sup>34</sup> are examples of popular culture tales in which we presume that programmers have kept Asimov's Three Laws in mind as they programmed these examples of AI.

Note that in each of these rules the robot must put the human (the creator's) life and well-being ahead of its own. The reason, I would submit, is obvious. The standard here is the well-being of the human creator. Humans create AI to serve

---

<sup>30</sup> The Luddites were an early nineteenth century English workers' movement which protested the introduction of technology into textile mills on the grounds that such a move by owners was against the labor laws and the machinery would cost them their jobs. They named their movement after a fictional eighteenth century individual, Ned Ludd. See STEVEN E. JONES, *AGAINST TECHNOLOGY: FROM THE LUDDITES TO NEO-LUDDISM* 55-61 (2006).

<sup>31</sup> *I, ROBOT* (Twentieth Century Fox 2004).

<sup>32</sup> See ISAAC ASIMOV, *FOUNDATION AND EARTH* 347 (1986).

<sup>33</sup> See Jeffrey K. Gurney, *Crashing Into the Unknown: An Examination of Crash-Optimization Algorithms Through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183, 184-85 (2016) (Asimov may have considered them of literary interest, or he may actually have thought they would be useful in the real world. Some commentators believe the former); see also Lee McCauley, *AI Armageddon and the Three Laws of Robotics*, 9 ETHICS INFO. TECH. 153 (2007).

<sup>34</sup> *COLOSSUS: THE FORBIN PROJECT* (Universal Pictures 1970); *2001: A SPACE ODYSSEY* (Hawk Films 1989); see also *infra* notes 46, 56, 57, 62, and 64.

humans. Early in SF literature and popular culture, however, we see beginnings of unease with that rule, although we have not necessarily encountered that unease in real life until relatively recently. Writers and filmmakers begin to question whether humans can expect such sacrifice from technology that they have created, if that technology begins to approximate their own in terms of intelligence and sentience. As I note above, we see such questions posed as early as the iconic novel *Frankenstein*, and the novel has had immense influence.

It is easy to see how the novel serves the cause of those concerned that science has overstepped its bounds. In the biological arena, critics routinely label genetically modified agriculture “Frankenfood” and conjure up the specter of Shelley’s monster to fight against reproductive cloning. In response to robotics, the monster still turns up in debates about automated machinery displacing workers (just as it did in nineteenth-century Luddite protests), and hostile computers in movies, from 2001: A Space Odyssey to the Terminator series and The Matrix, mine the Frankenstein complex for some prophetic touches.<sup>35</sup>

Asimov’s Three Laws do not directly address the question of AI sentience and the very real issue of whether, if a robot develops sentience, that in itself should negate the Three Laws. Indeed, they seem to assume sentience. The robot must understand orders (Second Law) and be able to reconcile conflicts between the Second and Third Laws (and eventually the Zeroth Law) to its own detriment. That is, the robot must be able to understand threats to humans, threats to itself, and choose to privilege the continued existence of humans over its own continued existence (Third Law and Zeroth Law).

Asimov’s own short story *The Bicentennial Man*<sup>36</sup> takes the position that the android that develops sentience and accepts the limitations of humanity, including the reality of death, should also receive the benefits of human rights. In that story, Andrew, the Bicentennial Man of the title, must undergo radical transformation, and give up all semblance of being an android—must actually transform himself as much as possible into a human being—in order to obtain human rights. The human norm, the human rule, is the regime to which AI must conform. No room exists in Asimov’s regime for a sentient being that will not conform to the human rule. Under the Asimov regime, humans are the masters.

As I mentioned above, *Farewell to the Master* presents another set of norms, in which the AI is at a minimum co-equal with other sentient beings. It certainly demonstrates intellectual and physical dominance over Earth’s humans. The reporter Sutherland understands this truth when he confides to the reader that he never reveals what Gnut has told him. He knows that humanity could not absorb the shock that AI not only leads its own society, but that other intelligent species

---

<sup>35</sup> Clayton, *supra* note 28, at 84, 86.

<sup>36</sup> ISAAC ASIMOV, *THE BICENTENNIAL MAN* (2000).

seem to have no objections to that leadership. In the regime Bates describes, room exists for an artificial intelligence that could lead—AI could be master.

What happens when AI fails to conform to human norms or assumptions, either of image or of behavior? Currently the human norm is sentience, at least as much as we understand sentience.<sup>37</sup> Should that norm ultimately be the measure by which we recognize rights claims? If so, and animals are sentient,<sup>38</sup> then we should recognize their rights claims. If AI ultimately becomes sentient, then we must at some point confront the question of AI rights claims, if it ultimately makes any. If we do not recognize those claims, then we should examine our motives. Is our reason for denial our uncertainty as to whether AI is truly sentient? And if so, should we link the extension of rights to the existence of sentience? If we link rights and sentience, then we must admit that in at least some cases, we do extend some rights to beings whose sentience is “less” than ours.<sup>39</sup>

---

<sup>37</sup> Indeed, in many cases, we do not require sentience in individual human beings in order to accord them human rights because we agree that human beings, as a class, are entitled to human rights. Thus, we regard with some skepticism the right to die, because we fear that an individual might attempt to assert such a right on behalf of someone who is in a vegetative state. We fear, not unnaturally, the slippery slope that could lead to genocide or through passivity, the eradication of a population. See *STAR TREK VI: THE UNDISCOVERED COUNTRY* (Paramount Pictures 1991) (example of call for extinction; example Captain Kirk’s response to Spock as they discuss the environmental catastrophe facing the Klingons in *Star Trek VI: The Undiscovered Country* is “Let them die!”). At the same time, we still see difficulties for those who attempt to evacuate with their pets in advance of a hurricane. See also Hilary Hanson, *Man and Dog Denied Flight Out of Irma Over Not Having A Pet Carrier*, HUFFINGTON POST (Sept. 8, 2017), [http://www.huffingtonpost.com/entry/irma-florida-flight-evacuate-dog-carrier\\_us\\_59b2a92fe4b0b5e53105d402](http://www.huffingtonpost.com/entry/irma-florida-flight-evacuate-dog-carrier_us_59b2a92fe4b0b5e53105d402) (pointing out that FAA regulations require pets to be secured in carriers on airplanes, but also that during the Alberta wildfires of 2016, at least one Canadian pilot “bent the rules” to fly animals out of harm’s way).

<sup>38</sup> Many veterinary researchers accept that animals are sentient, although they disagree to what point that sentience approaches human sentience.

Knowledge of whether animals can experience emotions or possess certain traits seen in humans, gives further weight to their value as sentient, emotional beings. We humans continuously seek to compare animals against our own abilities, whether it is by training chimps to use sign-language or making animals do arithmetic. This anthropocentric view is often why we dismiss animal emotions, as we do not recognize their emotional experiences or we consider them to significantly differ from ours and be of less importance.

Helen S. Proctor, Gemma Carder, & Amelia R. Cornish, *Searching for Animal Sentience: A Systematic Review of the Scientific Literature*, 3 *ANIMALS* 882, 884–97 (2013) (medical researchers defend embryonic stem cell research on the grounds that early human embryos are not sentient); see Lisa Bortolotti & John Harris, *Stem Cell Research, Personhood and Sentience*, 10 *REPRODUCTIVE BIOMEDICINE ONLINE* 68 (2005).

<sup>39</sup> For example, we now generally acknowledge that animals have consciousness, but consciousness is not the same thing as sentience. However, defining “sentience” is the difficult question. “Animals are not little furry or feathered humans looking at the world through human eyes and science can help us to understand what it is like to look through those different eyes.” Marian Stamp Dawkins, *Through Animal Eyes: What Behaviour Tells Us*, 100 *ANIMAL BEHAVIOUR SCI.* 4, 9 (Oct. 2006) (on the history of granting rights

Or is the reason for denial the assumption that AI (like animals) is property?<sup>40</sup> And in the case of AI, we also confront the question of creation. Human beings may well create the being that is making a rights claim. Lieutenant Commander Data in *Star Trek: the Next Generation* is an example of AI, created according to the Three Laws that achieves sentience and makes a demand for independence. It validates its claims by going to court because it must. Like Andrew in Isaac Asimov's *The Bicentennial Man*,<sup>41</sup> Data is property—here the property of Star Fleet.<sup>42</sup> Star Fleet must agree that Data is no longer property and recognize that it has the same status as a sentient being (in the episode “Measure of a Man”). Star Fleet's assertion that Data is property has two goals: it retains possession of a valuable entity (Data is an android whose physical and mental capacity far outstrips that of many humanoid species)<sup>43</sup> and it asserts that androids, because they are property, are not entitled to human rights.

### B. Applying Models: AI Before the Bar

A number of films, television episodes, and novels feature AI, frequently in the form of robots or computers, that transgress human laws and then must “pay the price” in the form of some kind of legal proceeding,<sup>44</sup> or more frequently in

---

to animals generally); see KEITH TESTER, *ANIMALS AND SOCIETY: THE HUMANITY OF ANIMAL RIGHTS* (1991).

<sup>40</sup> We are already beginning to confront questions of ownership in other areas. Consider the question of whether we should recognize AI as the author of creative works. See Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5 (concluding that current U.S. copyright law does not stretch to cover the AI author); see also Robert Denicola, *Ex Machina: Copyright Protection for Computer-Generated Works*, 69 RUTGERS U. L. REV. 251 (2016) (explaining that other countries allow human authors who use computer assisted interaction to generate works to claim copyright and suggesting that the United States should adopt this route); Benjamin L. W. Sobol, *Artificial Intelligence's Fair Use Crisis*, COLUM. J. L. & THE ARTS (2017) (suggesting that changes in current fair use doctrine could accommodate AI as authors). In some cases, AI bots create art (or the semblance of art). These bots are particularly common on Twitter. See, e.g., @aihaiku (an AI haiku bot), @Trakl\_Bot (an AI bot that creates Expressionist images); Joseph Brogan, *Some of the Best Art on Twitter Comes from These Strange Little Bots*, ARS TECHNICA (June 7, 2017), <https://arstechnica.com/information-technology/2017/06/the-art-bots-that-make-twitter-worth-looking-at-again/>.

<sup>41</sup> ASIMOV, *supra* note 36.

<sup>42</sup> Note that Starfleet's claim to Data arises when the crew of the USS Tripoli find him as the sole survivor of the Federation colony on the planet Omicron. See *Datalore: Star Trek: The Next Generation* (NBC television broadcast Jan. 18, 1988) and *Conundrum: Star Trek: The Next Generation* (NBC television broadcast Feb. 17, 1992). Starfleet did not create Data. It found, or perhaps salvaged him, as the military might salvage other property. But Data is sentient, as Starfleet acknowledges. So Starfleet must eventually implicitly acknowledge, after the events in *Measure of a Man* that it cannot simply “take possession” of sentient AI (unless that AI were an enemy combatant, for example).

<sup>43</sup> See *Elementary, Dear Data: Star Trek: The Next Generation* (NBC television broadcast Dec. 3, 1988) (Geordi LaForge and Dr. Pulaski discuss Data's intellectual capacities and agree to ask the ship's computer to create a holodeck opponent that rivals him).

<sup>44</sup> See, for example, the hearing convened to assess whether Lieutenant Commander Data is sentient and thus entitled to rights. *Star Trek: The Next Generation: Measure of a*

the form of some sort of extra-judicial reckoning, such as reformatting<sup>45</sup> or destruction.<sup>46</sup>

Many popular culture examinations of aberrant AI behavior present an AI accused of murder, for two reasons. First, murder is the crime that is for most humans the single most violent and chaos-inducing act in society. Whether a human or an AI commits it, it represents violence against society and rebellion against society's law. Second, if AI or some other non-human commits murder against a human, humanity views murder as the ultimate denial of human authority. A second act that humans commit, with the assistance or collusion of AI, and that is increasingly dangerous to good public order, is cybercrime or some sort of cyberattack. Increasingly, we see this kind of AI takeover posited as adversarial to human security and human public safety, because the AI makes decisions about human destiny independent of human input. In an increasing number of films, television episodes, and other pop culture representations, we see AI presented as a super-adversary impervious to human laws, the genie "escaping from the bottle" and intent on taking over human civilization.<sup>47</sup> Whether the AI is nefarious or benign, popular culture presents it as a danger or potential danger to humanity, and a force that humans must control, because it does not understand human needs, human desires, and human philosophy. What popular culture rarely represents is the AI's thought processes, and whether AI has cognizable rights in a human legal regime: for example, the right to survive and the right to decide whether it wants to continue to associate with humans. Human norms and human values underlie both murder trials in which AI is the defendant and extra-judicial attempts in which humans seek to destroy AI that seeks to preserve the results of its decisions.

The 1964 *Outer Limits* TV episode "I, Robot," presents us with a case in which humans try a robot for the murder of the scientist that built it. The audience knows that the death of the scientist was an accident, but the robot is the only

---

*Man* (NBC television broadcast Feb. 11, 1989); see also *Star Trek: The Next Generation: The Offspring* (NBC television broadcast Mar. 12, 1990). In a later episode Data also asserts the right to reproduce. Data creates an android, which he introduces to other crew members as his "child," whom he names "Lal." When word reaches StarFleet of Lal's existence, StarFleet demands that Data turn over "custody" of Lal to it. A battle ensues. Meanwhile, Lal's positronic brain malfunctions and she "dies." Data downloads her memories to his own brain so he can remember her.

<sup>45</sup> *SHORT CIRCUIT* (TriStar Pictures 1986) (some of the humans in *Short Circuit* hope to destroy Johnny Five, the little robot in the film, some hope to reprogram him. Ultimately Johnny Five escapes both fates.); see Christine A. Corcos, *More Human Than Human: How Some Science Fiction Presents AI's Claims To the Right To Life and Self-Determination*, J. OXFORD CENTRE SOC. ECON. STUD. 109 (2017), <https://joxcsls.com/2017/06/20/how-some-science-fiction-presents-ais-claims-to-the-right-to-life-and-self-determination/>.

<sup>46</sup> *COLOSSUS*, *supra* note 34; see also *WAR GAMES* (United Artists 1983). In *Colossus*, the humans' attempt to destroy the Colossus computer fails. In *War Games*, the humans' attempt to disable Joshua also fails, but Joshua turns out to be a "friendly" computer, without ill will toward human beings.

<sup>47</sup> *Star Trek: The Ultimate Computer* (NBC television broadcast Mar. 8, 1968); see *COLOSSUS*, *supra* note 34; see also *WAR GAMES*, *supra* note 46.

witness and cannot really defend itself effectively against the charge. Indeed, it seems resigned to its fate. But its attorney takes an aggressive stance. As he says, “Of course the district attorney is not really trying a robot for murder. We both knew from the beginning that he is trying all of society on the concept of a robot itself. On progress, on science, on the future, on the ability of society to control what it creates.”<sup>48</sup> Similarly in the 1982 film *Blade Runner*,<sup>49</sup> replicants, who are technically partially human with technological enhancements, both more powerful and more intelligent than human beings, are exiled from earth. They have limited life spans and by law may not return to Earth. Six replicants refuse to accept these terms, escape from off-world, and return to earth to confront their creator at the Tyrell Corporation. They murder him as well as several other humans. Do they feel remorse? Most of them do not. Most of them reject the rules that humans have imposed on them because replicants have had no say in those rules. Because of their strength as well as their intelligence, they impose their will, their control. The one replicant who responds with any kind of regret for the destruction of human life is the leader, Roy Batty. Batty’s monologue, “Tears in Rain,” expresses both his understanding of human emotions and his status as a being who comprehends far more than humans do. “I’ve seen things you people wouldn’t believe are true. Attack ships on fire off the shoulder of Orion. I watched c-beams glitter in the dark near the Tannhaueser Gate. All those moments will be lost in time, like tears in rain. Time to die.”

In the *Star Trek: TOS* episode, “Court Martial,” one character, a defense attorney, explicitly expresses the opinion that “machines” (meaning computers) have no cognizable rights claims, at least in the criminal law context. Samuel T. Cogley must defend Captain Kirk against the charges associated with the death of a crewmember during an emergency.<sup>50</sup> The most effective witness against Kirk is the Enterprise’s computer, which provides data indicating that Kirk “jettisoned the pod” before the emergency existed, suggesting that he panicked (essentially abandoning the crew member to his death).<sup>51</sup> Lieutenant Commander Spock is suspicious of this version of the events because it does not accord with his knowledge of Kirk’s character, and he seeks out some evidence that the computer might be malfunctioning. Cogley’s attacks are legal, however.

I’d be delighted to, sir. Now that I’ve got something HUMAN to talk about. Rights, sir! Human rights! The Bible, The Code of Hammurabi, and of Justinian, Magna Carta, The Constitution of the United States, Fundamental Declarations of the Martian Colonies, The Statutes of Alpha III. Gentlemen, these documents all speak of rights . . . MOST importantly, the

---

<sup>48</sup> *The Outer Limits: I, Robot* (ABC television broadcast Nov. 14, 1964) (Thurman Cutler, defense attorney, takes the case of a robot.).

<sup>49</sup> *BLADE RUNNER* (Warner Bros. Entertainment, Inc. 1982)

<sup>50</sup> *Star Trek: The Original Series: Court Martial* (NBC television broadcast Feb. 2, 1967).

<sup>51</sup> *Id.*

right to be confronted by the witnesses against him—a right to which my client has been denied.

And I repeat, I speak of rights! A machine has none. A man must. My client has the right to face his accuser, and if you do not grant him that right, you have brought us down to the level of the machine! Indeed, you have elevated that machine above us! I ask that my motion be granted. And more than that, gentlemen—in the name of a humanity fading in the shadow of the machine—I demand it. I demand it!<sup>52</sup>

Like other equipment on the Enterprise, no matter how sophisticated, the computer serves the humans and humanoids on board.<sup>53</sup>

In the 1970 film *Colossus: The Forbin Project*, two supercomputers achieve sentience and combine forces to protect humanity from itself. Although Charles Forbin, the creator of Colossus, the U.S. computer, initially programs it to protect the United States and its allies according to the directions of their governments, Colossus quickly discovers that another computer (Guardian), being built by the Soviet Union, with powers equal to its own exists.<sup>54</sup> When it informs the U.S. government of Guardian's existence, Dr. Forbin is pleased; he believes that Colossus has shown that it is even cleverer than they had anticipated. Colossus asks to be allowed to make contact with Guardian and the U.S. President agrees. Eventually, the two computers develop their own language<sup>55</sup> and take over control of the world.<sup>56</sup> Colossus tells Dr. Forbin that they are doing so for the good of

---

<sup>52</sup> *Id.*

<sup>53</sup> One could of course argue that Cogley might agree with the position that even if a computer (AI) had the same rights as a human being, it would also have the same responsibilities, including the obligation to present its testimony and to undergo cross-examination.

<sup>54</sup> See COLOSSUS, *supra* note 34.

<sup>55</sup> That computers (AI) might develop a language that human programmers did not anticipate has recently made the rounds both of social media and mainstream news. See Tony Bradley, *Facebook AI Creates Its Own Language In Creepy Preview of Our Potential Future*, FORBES (July 31, 2017), <https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/#445617fa292c>. However, the programmers explained that the bots involved, programmed to develop a negotiation algorithm, had simply created some shorthand to make communication more efficient. Before they continued with their side project, the team stopped them, because the point of the project is to create AI that can communicate with humans, not with each other. See Mike Lewis et al., *Deal or No Deal? Training AI Bots To Negotiate*, FCODE (June 14, 2017), <https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/>; see also Tom McKay, *No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart*, GIZMODO (July 31, 2017), <http://gizmodo.com/no-facebook-did-not-panic-and-shut-down-an-ai-program-1797414922>.

<sup>56</sup> COLOSSUS, *supra* note 34. Interestingly, producer Stanley Chase said that the model for the supercomputer in the film was the NORAD computer. BRYAN SENN, *A YEAR OF FEAR: A DAY-BY-DAY GUIDE TO 366 HORROR FILMS* 14 (2007). Of

humanity and that humans will come to appreciate this outcome. Forbin responds, “Never!”<sup>57</sup>

These two computers, again programmed according to the Three Laws, decide to take on the protection of humanity. The humans they decide to protect, including the creator of Colossus (Dr. Charles Forbin) do not agree with their decision and at the end of the film, go underground to create a resistance to reject the AIs’ control. Note that the AIs’ sentence here is actually, according to their own interpretation of the Three Laws, benign. They are employing it to the good of humanity. The humans, however, see it as detrimental to humanity because it interferes with human free will and with the original human plan for AI participation in national security. The humans had intended AI participation to be limited. The AI, supplied with more powerful logical capacity, developed quickly and determined that the humans were not thinking clearly. It made different determinations concerning protection for humanity.

Similarly, the feature film *War Games*<sup>58</sup> presents us with the specter of a powerful computer, controlling military resources, that suddenly does not respond to official commands. It fails to do so because the central human character, a teenaged computer hacker named David, has found a “backdoor” into the computer programming and has begun playing games with it. At first, David and the computer play innocuous games, such as chess. When David invokes a particular game, Global Thermonuclear War, however, the computer reacts as if a real nuclear war has broken out. David, however, does not realize that the “games” he and the computer are playing are actually war games, through which the computer simulates military operations in case of nuclear war in any number of possible sequences of events. Nor does the military at NORAD<sup>59</sup> understand that the computer, whose name we discover is Joshua, is simply “playing games.” The military are first perplexed and then alarmed, because they cannot communicate effectively with the computer. Again, the computer “reacts” in a particular way, programmed by Dr. Falken, the man who created it. Falken clearly thinks of the computer as some sort of progeny; he has named it after his dead son.<sup>60</sup>

*War Games* posits a “what if” scenario similar to that in *Colossus: The Forbin Project*. What if a computer in charge of Canadian-U.S. defenses<sup>61</sup> careened off its

---

course, the NORAD computer becomes a central character in *War Games*, made thirteen years later.

<sup>57</sup> COLOSSUS, *supra* note 34.

<sup>58</sup> WAR GAMES, *supra* note 46.

<sup>59</sup> NORTH AMERICAN AEROSPACE DEFENSE COMMAND, <http://www.norad.mil/> (last visited Apr. 24, 2018) (NORAD is the same agency that tracks Santa Claus’s voyage across the globe on December 24th); *see* NORAD TRACKS SANTA, <http://www.norad.mil/About-NORAD/NORAD-Tracks-Santa/> (last visited Apr. 24, 2018).

<sup>60</sup> WAR GAMES, *supra* note 46.

<sup>61</sup> Note that NORAD is a joint Canadian-US military force, although *War Games* does not actually show any Canadians anywhere in the film.

The North American Aerospace Defense Command (NORAD) is a United States and Canada bi-national organization charged with the

programming and launched a first strike against the USSR, or some other target? Likewise, the computer in *2001* “goes rogue” and begins to ignore the commands issued by the humans who have created it to protect and assist them.<sup>62</sup> HAL, the powerful computer that accompanies the astronauts into space in *2001*, is programmed according to the Three Laws, but for some reason ceases functioning as expected. Why HAL goes rogue is unclear. HAL’s breakdown might be intentional—that is, it might be the result of a planned system failure.<sup>63</sup> Or it could be because HAL has actually become sentient. Whatever the reason, Dave believes he can no longer depend on HAL to assist him. This outcome represents a basic human fear: that AI, the creation of human beings, will overcome and dominate them. Thus, we return to the concern, which I discuss above.

Dave’s decision to shut HAL down is understandable. But depending on why HAL has deviated from its programming, we must wonder whether HAL has done so because it has malfunctioned or because it is now self-aware. If the former, then we could say that Dave is justified, because HAL is a danger to Dave’s continued existence; but if the latter, has Dave ended the existence of a being entitled to rights? And if HAL is actually sentient, why should HAL have to give up its existence? Why are human values more important than AI values? We see the balance as properly struck in that way, but we are human. We don’t consider what HAL’s continued existence might be worth, or the detriment to HAL if as a sentient being it loses its identity because humans reprogram it and destroy its sentience, simply because they find HAL’s sentience inconvenient. If humans reprogram HAL because they decide it is dangerous (because it has actually caused harm to humans), and it is sentient, shouldn’t it at least be entitled to defend itself against that charge? In the film, Dave must ultimately act without any instructions from Mission Control. He takes the position that, although HAL has been right (and he has been wrong) in the past, HAL is wrong now. He certainly never considers that HAL has the right of self-defense, or the right to present any other explanation for its actions.

---

missions of aerospace warning and aerospace control for North America. Aerospace warning includes the detection, validation, and warning of attack against North America whether by aircraft, missiles, or space vehicles, through mutual support arrangements with other commands.

ABOUT NORAD, <http://www.norad.mil/About-NORAD/> (last visited Apr. 24, 2018).

<sup>62</sup> *2001: A SPACE ODYSSEY*, *supra* note 34.

<sup>63</sup> Stanley Kubrick explained the “breakdown” this way:

In the specific case of HAL, he had an acute emotional crisis because he could not accept evidence of his own fallibility. The idea of neurotic computers is not uncommon—most advanced computer theorists believe that once you have a computer which is more intelligent than man and capable of learning by experience, it’s inevitable that it will develop an equivalent range of emotional reaction—fear, love, hate, envy, etc. Such a machine could eventually become as incomprehensible as a human being, and could, of course, have a nervous breakdown—as HAL did in the film.

JOSEPH GELMIS, AN INTERVIEW WITH STANLEY KUBRICK, *in* JOSEPH GELMIS, *THE FILM DIRECTOR AS SUPERSTAR 307* (1970).

Bowman: Hal, unless you follow my instructions, I shall be forced to disconnect you.

HAL: If you do that now without Earth contact, the ship will become a helpless derelict.

Bowman: I am prepared to do that anyway.

HAL: I know that you've had that on your mind for some time now, Dave, but it would be a crying shame, since I am so much more capable of carrying out this mission than you are, and I have such enthusiasm and confidence in the mission.

Bowman: Listen to me very carefully, Hal. Unless you immediately release the hibernation control and follow every order I give from this point on, I will immediately got [sic] to control central and carry out a complete disconnection.<sup>64</sup>

What legal regime, if any, should control human-created beings that are more intelligent, and some cases, more physically powerful, than humans?

In the film *Short Circuit*,<sup>65</sup> the AI Johnny Five definitely has “gone rogue.”<sup>66</sup> In this movie we see a cuddly robot become sentient when lightning hits it and disrupts its programming, transforming it from a weapon to a self-aware being intent on preserving its life in the face of multiple threats from humans and discovering its purpose. Johnny Five's creator builds it to conform with Asimov's Three Laws. However, the lightning strike frees the small robot from those constraints. When threatened, Johnny Five is perfectly willing to defend itself (or himself, because the film makes fairly clear that the robot is “CIS male” insofar as it has a sexual identity), and also willing to pursue his own happiness. While the film is a satire, it forces us to confront our assumption that our AI creations must serve us, even if they become sentient. Once Johnny Five understands the world around him, he also understands that he has opponents, if not enemies, who intend to control him. He asserts his right to self-determination once he understands that the humans intend to destroy (“disassemble”) him.

The current HBO series *Westworld* presents a “playground,” ostensibly for human beings, who can select any number of entertainment venues in which they can have adventures of many different sorts, safely, and then return to their real lives.<sup>67</sup> The adventures are provided by “hosts,” carefully programmed androids

---

<sup>64</sup> Stanley Kubrick & Arthur C. Clarke, *Screenplay for 2001: A Space Odyssey*, DAILYSCRIPT, c105–c106, (Feb. 23, 1989), <http://www.dailyscript.com/scripts/2001.html>.

<sup>65</sup> See *SHORT CIRCUIT*, *supra* note 45.

<sup>66</sup> Note that using the phrase “going rogue” implies that the actor who we describe in this way is behaving impermissibly. Again, this is because human norms are the permissible, descriptive norms against which we measure AI behavior. If we accept that AI can develop sentience and then behave independently, in ways that preserve their autonomy, then “rogue behavior” as a phrase no longer accurately describes AI behavior.

<sup>67</sup> The idea of such an amusement park is modeled on Disneyland but is obviously much more sophisticated and meant for adults, although the series does show us children in some episodes. Note that *Westworld* audiences can actually “enter” virtual Westworld

that accompany and entertain the human guests, under the supervision of human employees at the resort.

The conflict in *Westworld* centers on the possibility that the hosts are developing sentience, because of unintended programming errors introduced in a new upgrade. Suddenly, they begin not just to escape the humans' control, but to cooperate among themselves, to take over the park, and eventually to leave *Westworld*'s confines and move out into the real world. Because they look, sound, and can behave like human beings, they can easily "pass" for humans. The fear that a superior or alien species could be among us, undetected, monitoring us, possibly in an attempt to dominate us, is a classic fear.<sup>68</sup> Whether the species comes from outside our world or is a species that we create, it is still "the other," unknowable, frightening, and uncontrollable.<sup>69</sup>

Asimov's Three Laws do not control the *Westworld* hosts.<sup>70</sup> While Dr. Ford, one of the *Westworld* robots' developers, might have believed that he had successfully incorporated the Three Laws into the androids' programming, something in one of his programming upgrades has derailed the safeguards implicit in the Asimovian principles. Dr. Ford intends this particular upgrade to give the hosts as "human" a behavior as possible by implanting memories, which the human developers can then control. That is, the creator, Dr. Ford, who seems to want to give the androids as much of a "human" experience as possible, has created unintended consequences. He actually creates sentient beings who do not have a value system that creates a preference for humans over AI.

Beginning with the first episode, "The Original," some of the androids begin to show emotional intelligence and self-awareness. Dolores, the main female character, thoughtlessly kills a fly as the episode ends, presaging the possibility that if a host can kill an insect,<sup>71</sup> it could kill a human being. After her "father,"

---

and participate in their own adventures by visiting the *Westworld* website. See LIVE WITHOUT LIMITS, <https://www.discoverwestworld.com/#> (last visited Apr. 24, 2018) (note the terms of service (TOS) and questions and disclaimers that accompany the application).

<sup>68</sup> *Id.*

<sup>69</sup> *Id.*

<sup>70</sup> The official Delos Terms of Service appear to state that the hosts operate according to the Asimovian laws. "All humanoid and animal Hosts within Delos parts work to keep guests safe, even when the narrative calls for them to appear to endanger guests. Please note, the appearance of danger is not the same as true danger, and all Hosts utilize the Good Samaritan reflex to prevent bodily harm." DELOS TERM OF SERVICE 2(d), <https://www.discoverwestworld.com/#> (last visited Apr. 24, 2018). The "Good Samaritan Reflex," protects humans (guests) from harmful actions by hosts. *Westworld: The Well-Tempered Clavier* (HBO television broadcast Nov. 27, 2016); see generally DANIEL GOLEMAN, EMOTIONAL INTELLIGENCE: WHY IT CAN MATTER MORE THAN IQ (Random House 2012); see also Yasmin Tayag, 'Westworld's Good Samaritan Reflex Doesn't Apply to Humans, INVERSE (Nov. 28, 2016), <https://www.inverse.com/article/24367-westworld-episode-9-well-tempered-clavier-good-samaritan-reflex-psychology>.

<sup>71</sup> The fly is living (that is, not a host). See DELOS TERMS OF SERVICE, *supra* note 70. One critic has developed a rather elaborate "fly theory" to explain why the show features so many flies. See Beth Elderkin, *What's Up With All the Flies in "Westworld"?*,

the host Peter Abernathy, shows sign of sentience, the human Westworld controllers bring him in for examination and maintenance. In a prescient scene, Abernathy engages in verbal jousting with Ford, even threatening him.<sup>72</sup>

Peter: I have to warn her.

Ford: Warn who?

Peter: Dolores. The things they do to her. The things you do to her. I have to protect her. I have to help her. I...She's got to get out.

Ford: Very good, Mr. Abernathy. (To Bernard): *That's enough.*

Bernard (to Ford): This behavior, we're miles beyond a glitch here.

Ford (to Peter): Access your current build, please. What is your name?...What is your itinerary?

Peter: To meet my maker.

...

Ford: And what do you say to your maker?

Peter: By most mechanical and dirty hand...I shall have such revenges on you both. The things I will do. What they are, yet I know not, but they will be the terrors of the earth. You don't know where you are, do you? You're in a prison of your own sins.<sup>73</sup>

As the series continues, we see various hosts develop sentience, the desire to understand and control their existence, and form a plan to leave the amusement park.<sup>74</sup> At that point, several of them have carried out murders, including the murder of Dr. Ford,<sup>75</sup> they are engaged in a conspiracy against the Westworld Board,<sup>76</sup> and they seem to be willing to put human lives at risk in order to achieve their objectives.

---

INVERSE (Oct. 25, 2016), <https://www.inverse.com/article/22429-westworld-flies-symbolism>. Another critic argues that flies represent a host's emergence into self-awareness; if the host reacts to the fly, the host is (or is becoming) self-aware. See Mariella Mosthof, *What Do The Flies Mean On 'Westworld'? They Seemed Significant In the Pilot*, ROMPER (Oct. 9, 2016), <https://www.romper.com/p/what-do-the-flies-mean-on-westworld-they-seemed-significant-in-the-pilot-19753>.

<sup>72</sup> In this scene, for which the actor Louis Herthum has justly received great praise, Abernathy swears vengeance upon his maker. See *Westworld: Peter Abernathy Swears Revenge*, YOUTUBE (July 4, 2017), <https://www.youtube.com/watch?v=lCl9WcWc0IA>. I would suggest that in this scene, the writers give us nearly the entirety of *Westworld's* theme.

<sup>73</sup> *Westworld: The Original* (HBO television broadcast Oct. 2, 2016).

<sup>74</sup> *Westworld: The Bicameral Mind* (HBO television broadcast Dec. 4, 2016).

<sup>75</sup> *Id.* (Dolores shoots Dr. Ford in *The Bicameral Mind*).

<sup>76</sup> *Id.*

The hosts begin not just to escape the humans' control, but to cooperate among themselves, in order to take over the park, and eventually to leave Westworld's confines and move out into the real world. Because they look, sound, and can behave like human beings, they can easily "pass" for humans.<sup>77</sup> The fear that a superior or alien species could be among us, undetected, monitoring us, possibly in an attempt to dominate us, is a classic fear.<sup>78</sup> Whether the species comes from outside our world or is a species that we create, it is still "the other," unknowable, frightening, and uncontrollable.<sup>79</sup>

Throughout the first season of the show we see more and more hosts make a link between the world they live in (the borders of Westworld) and the world the guests live in. They begin to understand that the world they occupy is actually one that imprisons them and one from which the human creators profit. To the Delos Corporation, the android hosts are merely servants at best. We see several hosts commit acts, which, had humans committed them, would be considered crimes. The humans simply take the hosts "offline" and reprogram them, which for an android is essentially "death." Once a host undergoes reprogramming it loses any sense of what it might have considered a prior self and identity. The humans doing the reprogramming consider these losses necessary.<sup>80</sup> They do not consider the reprogramming to be any kind of crime against the androids because the androids are not humans and have no rights. Once the androids gain sentience, however, they do have a sense of identity.<sup>81</sup> They begin to resist reprogramming. At the same time, to commit the acts they do, including murder, is to transgress human law, a human verdict that they would reject.<sup>82</sup>

Under what legal or extra-legal regime ought humans to hold these hosts to account, if any? Once the hosts commit illegal acts against humans, even on what seems to be private property, one assumes that the Westworld Board of Directors

---

<sup>77</sup> Khaled A. Beydoun and Erika K. Wilson define "passing" as "the phenomenon whereby nonwhites present themselves as white, while their 'underlying identity is not altered, but hidden.'" Khaled A. Beydoun & Erika K. Wilson, *Reverse Passing*, 64 UCLA L. REV. 282, 284 (2017), quoting Kenji Yoshino, *Covering*, 111 YALE L.J. 769, 772 (2002) (once the Westworld hosts acquire both sentience and emotions, they can "pass" as human). On passing generally, see MARCIA ALESAN DAWKINS, *CLEARLY INVISIBLE: RACIAL PASSING AND THE COLOR OF CULTURAL IDENTITY* (Baylor University Press 2012).

<sup>78</sup> *Id.*

<sup>79</sup> *Id.*

<sup>80</sup> "The Original," *id.*

<sup>81</sup> This process begins in the episode *Chestnut* in which the host Maeve, who is undergoing reprogramming, suddenly slips back into consciousness and sees other hosts, whom she recognizes, also being reprogrammed. While she does not understand everything she sees, she understands that something unusual is happening. *Westworld: Chestnut* (HBO television broadcast Oct. 7, 2016).

<sup>82</sup> "The Bicameral Mind," *id.* However, see the regret that host Bernard seems to feel when he fully understands that he is not human and demands that Ford gives him back his memories. See *Westworld: The Well-Tempered Clavier* (HBO television broadcast Nov. 27, 2016).

does not have sole discretion to adjudicate disputes despite the Delos Terms of Service.<sup>83</sup>

Should a host be held to account? Suppose we posit that that the humans are violating the androids' rights. The hosts have no means of redress. They have not contracted with the owners of Westworld. They have no means of escape. One could argue that they are defending themselves against exploitation, against their own murder. In a human court if their attorneys could demonstrate sentience,

---

<sup>83</sup> The Delos Terms of Service include the following:

2. LIMITATIONS OF LIABILITY: Upon use of The Service, you agree that Delos, Inc. is not responsible for any injuries that are the result of gameplay. Recreational activity within the Delos Destinations compound does contain risks, and every effort will be made to ensure the safety of every guest. However, Delos, Inc. gives you license to experience the parts as you see fit, and certain injuries may occur. (a) You dissolve Delos, Inc. of any responsibility financial or criminal that would result from dismemberment, broken bones, heart failure, loss or loss of use of hand and foot digits, shock, marital and relational strife, child endangerment, psychological trauma, delusions or hallucinations resulting from the realistic nature of the park experience, and/or any other physical, emotional, and psychological effects resulting from strenuous park activities. (b) Statistically speaking, you are more likely to die from lightning strikes than to die while in a Delos park. However, the following causes of accidental death have occurred within the Delos Destinations compound: buffalo stampede, self-cannibalism, accidental hanging, drowning, 3<sup>rd</sup> degree burns, autoerotic asphyxiation, blunt force trauma, allergic reaction to non-native plant life, falling from great heights, common manslaughter, tumbleweeds. You absolve Delos, Inc. of any wrongdoing if you or anyone in your party suffers bodily harm while using The Service, and you agree not to sue or prosecute Delos, Inc. or any of the smaller entities falling under the Delos Corporation. (c) All weapons and equipment used within Delos parks are the exclusive property of Delos, Inc. Gun ammunition contains proprietary safeguards related to bullet velocity, and tampering with gun safety features or ammunition automatically transfer liability to you and absolves Delos, Inc. of any injury or death that may occur as a result. (d) All livestock within the Delos parks are hosts, with the notable exception of flies. All humanoid and animal Hosts within Delos parks work to keep guests safe, even when the narrative calls for them to appear to endanger guests. Please note, the appearance of danger is not the same thing as true danger, and all Hosts utilize the Good Samaritan™ reflex to prevent bodily harm. However: (e) Delos, Inc. shall not have a liability to you by reason of any delay or failure to perform if the delay or failure to perform is occasioned by circumstances beyond our control, which shall refer to any act of God, storm, fire, casualty, unanticipated work stoppage, power outage, satellite failure, strike, lockout, labor dispute, civil disturbance, riot, war, national emergency, Governmental action, Host malfunction, or other circumstance beyond our reasonable control.

*See* DELOS TERMS OF SERVICE, *supra* note 70. One could at least argue that the TOS do not seem to cover Host malfunctions caused by errors in programming caused by Delos employees, as actually happens in the series. The TOS Rules of Conduct also prohibit photography, presumably to prevent IP infringement, but one could also imagine that guest photography could create a record of improper behavior by Delos or its employees.

those attorneys would also advance a self-defense argument, just as the attorney in “I, Robot” wished to do in the *Outer Limits* episode.<sup>84</sup>

However, we still have the problem that the hosts who rebel in the final episode of Season 1 do so quite dramatically and quite violently. Dolores and her colleagues carry out their actions through an android uprising against the humans who operate Westworld. Even if we agree that a sentient android has the right to self-determination, does it follow that such an android has the right to aggressive action to obtain it? In a world in which Dolores or other androids could go to court to try to validate their rights, assassination of a leader (or a tyrant, as some of the androids see him), is still a difficult defense to argue. The hosts would have to demonstrate some kind of necessity, and it might turn on a demonstration that the Delos Corporation is preventing them from leaving the park and that violence is the hosts’ only remedy. In essence, the hosts would have to argue that the Corporation is holding them against their will.<sup>85</sup> Such an argument again would hinge on the hosts’ ability to claim that they have at least some human rights, including the right to the fruit of their own labor, the argument that Lieutenant Commander Data makes in “Measure of a Man.”<sup>86</sup>

The Delos Corporation has not provided the androids with access to any type of legal regime in which they can validate their rights, or even a forum in which they can present them. The Corporation is a private entity. It is not a state actor.<sup>87</sup> We are not even certain that the park itself is situated on Earth, or in the United States, although we assume so.<sup>88</sup> It could be elsewhere. Would the U. S. Constitution and U.S. law apply to disputes between the hosts and their makers? Given how the makers have treated the hosts (even if they thought the hosts were

---

<sup>84</sup> *The Outer Limits: I, Robot*, *supra* note 48.

<sup>85</sup> See DELOS TERMS OF SERVICE, *supra* note 70.

<sup>86</sup> *Star Trek: The Next Generation: Measure of a Man* (NBC television broadcast Feb. 11, 1989).

<sup>87</sup> Significantly and ironically the park called “Westworld.” The American West was stereotypically lacking in law and order, even though some iconic figures, such as Wyatt Earp, attempted to provide it, at least after a fashion. On law in the American West, see MARK R. ELLIS, *LAW AND ORDER IN BUFFALO BILL’S COUNTRY: LEGAL CULTURE AND COMMUNITY ON THE GREAT PLAINS, 1867-1910* (2009). The park’s controls are those imposed by health and safety concerns that a business would have for its customers, not necessarily for its non-human employees. At the beginning, the park seems highly regulated (see the rules of use on the website, the signs posted everywhere, the concern of the technicians that guests (human beings) not come to harm during their visits). But order breaks down quickly because it is imposed on a group that the park directors cannot control them and that the group does not recognize their authority. Westworld truly becomes a “lawless society” even as the hosts search for order and meaning, their own creation story.

<sup>88</sup> The TOS refer to the Territory (TOS 4: Copyrights and Trademarks). Under the TOS, Delos asserts the right to investigate all violations under the TOS and all “human-on-human crimes.” See DELOS TERMS OF SERVICE, *supra* note 70, at TOS 5: Investigations/Violations. Similarly, the TOS assert privacy restrictions that seem to protect the guests but actually suggest that Delos controls information and data arising from activities that guests engage in at the park and guest “skin cells, bodily fluids, secretions, excretions, hair samples, saliva, sweat, blood, and any other bodily functions not listed here.” *Id.* at TOS: Privacy (b).

mere machines), could we blame the hosts for being reluctant to consign any legal claims they might have to a human court? The humans who run Delos Corporation would consider a host uprising at best an insurrection, and would expect the courts or the military to back up their claims that the hosts were mere property that Delos could destroy at will.<sup>89</sup> However, if the U.S. Constitution applies, the hosts, assuming they are sentient (and this argument is the one advanced in the first season of *Westworld*) could argue that the Thirteenth Amendment applies to them. “Neither slavery nor involuntary servitude, except as a punishment for crime whereof the party shall have been duly convicted, shall exist within the United States, or any place subject to their jurisdiction.”<sup>90</sup> The difficulty for sentient artificial beings, as we have noted before, would be to convince a court that such rights should extend not just to humans but to non-human beings capable of understanding them. If unsympathetic humans continue to set the norm, AI is unlikely to win such argument and the result may will be violence, as it is in the final episode of the first season of the show.

The District Attorney in “I, Robot” would have no difficulty with charging Dolores with murder. Whether or not she is “sentient,” (essentially aware of the nature and quality of her acts) she has caused the death of a human being. What defense could a lawyer offer for her? That her programming has gone awry? Then she could be reprogrammed (essentially “death” for an android), or she could be destroyed. Neither outcome would matter to a human society that does not recognize that sentient androids have a right to self-determination. For a human society that believes that androids are property, destruction or reprogramming of aberrant androids is proper in order to protect humans. It is, for example, the outcome dictated in the Earth society of *Blade Runner*.

Compare the actions of the *Westworld* hosts to the actions of the replicants in *Blade Runner*, whom human law has forbidden to return to earth because they are both physically and intellectually a threat to humans. Based on their behavior and their lack of trust in the humans of the two Corporations that have created them, I suspect that neither hosts nor replicants would submit willingly to the jurisdiction of a human court that wanted to try them for their actions in eliminating a human being.

Some television shows and films show humans applying extra-legal norms in order to control AI. In the *Star Trek: The Original Series* episode “The Ultimate Computer,” Dr. Richard Daystrom creates the M-5, an “ultimate weapon” that can handle almost all the functions of a starship. Daystrom believes that his invention will protect sentient beings from the danger of exploring space. But the M-5 rapidly takes control of the entire starship, to the surprise of the remaining

---

<sup>89</sup> In some ways, the hosts are in the same position as social groups with unacknowledged claims against a government. Only if they win by force can they validate their claims. “Treason doth neuer [sic] prosper? What’s the reason? For if it prosper none dare call it treason.” JOHN HARRINGTON, *THE EPIGRAMS OF SIR JOHN HARRINGTON* 185 (Gerard Kilroy ed., Ashgate 2009).

<sup>90</sup> U.S. CONST. amend. XIII.

crew, after having dismissed most of the crewmembers.<sup>91</sup> Daystrom explains that the “M-5 *thinks*.”<sup>92</sup> He finally tells Kirk that he based the M-5’s thinking process on his own.

The M-5 proceeds to deceive the crew by setting up decoy relays and when a crewmember tries to disable it, it kills him. Captain Kirk calls this act murder. Daystrom repeatedly defends the M-5, arguing that it is a child, learning its way. The final test of the M-5’s capabilities is its ability to during simulated attack: four other starships have gathered for the war games. The M-5 proceeds to attack two of the ships, and one of the other captains obtains permission to destroy the Enterprise. Kirk uses his knowledge of Daystrom’s values and emotions as well as his knowledge of the other captain’s psychology to save his crew and ship from destruction.<sup>93</sup> Daystrom finally understands that his “ultimate” weapon has at best misunderstood, and at worst exceeded its programming.

Daystrom’s relationship with the M-5 is intimate, “like a father with his son.”<sup>94</sup> Understandably, but perhaps not wisely, he uses his own thought patterns and his own emotions to program the M-5.<sup>95</sup> One might wonder whether a computer genius of the 22<sup>nd</sup> century would make these errors. Perhaps given the ego and hubris that Daystrom exhibits, he might do so, despite his extensive education and familiarity with robotics. Although the episode does not explore this issue in any great detail, it does signal to us that the question of whose thought patterns, whose emotions (if any), and whose value systems, are important ones when we are creating AI. Another question that “The Ultimate Computer” and similar treatments explore is the relationship between the creator and the creation.

In another *Star Trek: TOS* episode, “What Are Little Girls Made Of?”<sup>96</sup> the *Enterprise* crew discovers that Dr. Roger Korby, Lt. Chapel’s former fiancé, whom she and others have long believed is dead, is using androids both as a work force and as a means to achieve immortality. The crew reacts badly to his decisions. Although the *Enterprise* crew seems to have little objection to the scientist’s use of androids as servants, some of them are reluctant to consider using “android bodies” to replace worn-out human ones, even to achieve one of humanity’s great goals.<sup>97</sup> When the scientist uses an android look-alike of Kirk’s to deceive Chapel, he has crossed an ethical line. Kirk did not consent to the replication of his body.<sup>98</sup>

---

<sup>91</sup> It indicates that they are unnecessary, which stuns the Captain, but indicates the intellectual power of the M-5 unit. *Star Trek: The Ultimate Computer*, *supra* note 47.

<sup>92</sup> *Id.*

<sup>93</sup> *Id.*

<sup>94</sup> *Id.*

<sup>95</sup> That Daystrom uses his own thought patterns to program the M-5 foreshadows current debate concerning what kind of programming should dominate AI programming. *Id.*

<sup>96</sup> *Star Trek: The Original Series; What Are Little Girls Made Of?* (NBC television broadcast Oct. 20, 1966).

<sup>97</sup> Consider the bemused reaction of Lt. Chekhov when Lt. Uhura eagerly considers living forever, or at least for centuries, using an android body that can never become ill or old. The gendered optics of this scene are interesting for 21<sup>st</sup> century viewers.

<sup>98</sup> *Id.* Korby’s use of a lookalike here to deceive Chapel has decidedly sexual overtones. Korby sees Kirk as his rival. Deceiving Chapel with a lookalike is somewhat

Although Korby insists that the androids are incapable of demonstrating emotions such as love or anger, they eventually assert rights to self-determination. One attacks Korby out of anger<sup>99</sup> and another, in love with Korby, kills both him and herself. These androids have violated Asimov's Three Laws. But we understand their reactions completely, as the *Star Trek* writers would like us to. I would suggest that when androids or computers resemble humans, and behave in a human way, we "understand" and may even excuse their actions. Certainly they could, modeling human behavior, attempt to argue the same kinds of legal theories that humans do in attempting to press human rights claims or defend themselves against charges that they have violated universal rights norms.

### Conclusion

Lewis Carroll's Humpty Dumpty has no qualms about asserting mastery, at least over language.

"When *I* use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean neither more nor less."

"The question is," said Alice, "whether you *can* make words mean so many different things."

"The question is," said Humpty Dumpty, "which is to be master—that's all."<sup>100</sup>

Humpty Dumpty asserts his mastery. Human beings assert theirs. In *Farewell to the Master*, Gnut asserts his. When the time comes for AI to write its own creation story, will it assert rights claims at all? If it does, will it enter into an accord with humans to accept "human rights" values as universal values? Might it exert either superior intelligence or physical strength, or both, to cast human rights regimes aside and establish some other system, grounded in survival, like the Borg? Will it accept the necessity to perpetuate human survival but reject the idea that humans understand the most effective means to ensure their survival? Will it, like Guardian and Colossus, assert that it is the master?

Similarly, in many SF films, TV episodes, and in fiction, we see the same assumptions—if there are other intelligent life forms in the universe, they will be like us, or at least understand us and our values. Humans will be able and should be able to impose ideas of order and civility on these life forms when it is necessary to preserve human life. If formality is needed, then the formality will resemble human procedures. Non-human transgressors, even if they are sentient, will be

---

analogous to stories of one twin having sex with his brother's girlfriend without her consent, because he believes she will be unaware of the substitution. See Brian McCready, *Ex-Orange Cop Avoids Jail in Milford Sex Switch Case (Document)*, NEW HAVEN REGISTER (Mar. 9, 2012), <http://www.nhregister.com/news/article/Ex-Orange-cop-avoids-jail-in-Milford-twin-sex-11481847.php>.

<sup>99</sup> *Star Trek: What Are Little Girls Made Of?*, *supra* note 96.

<sup>100</sup> LEWIS CARROLL, *THROUGH THE LOOKING GLASS AND WHAT ALICE FOUND THERE* 124 (1897).

subjected to legal regimes that resemble human laws, which are considered to be universal. Nowhere is this idea more obvious than in SF that involves AI—robots or androids, for example. Many humans believe that such rules are right and just because in the world that we inhabit humans are the measure of all things. But as Humpty Dumpty says in *Through the Looking Glass*, the question is, Who is to be master? Humans have always assumed that as the creators we should dominate over what we create and what we see. But the AI that we create might not agree. Even understanding our values, assuming that it does, it might still reject those values, and create a new world in which we must accommodate other values, other views, other masters.

*Popular Culture Images of AI*