When Robots Choose to Die: A Survey of Robot Suicide in Science Fiction

Liz W Faber, PhD
Lecturer in Academic Writing & Composition
Manhattanville College
2900 Purchase Street
Purchase, New York 10577
914-323-5268
E-mail: Elizabeth.Faber@mville.edu

A version of this paper was presented at the PCA/ACA National Conference
Washington, D.C.
April 18, 2019

In 2007, GM ran an ad featuring a robot that loses its job and subsequently destroys itself by jumping off a bridge. Although the ad ended in the robot jolting awake to find that its death had been a nightmare, public outrage regarding the flippant treatment of suicide led GM to revise the ad a week later (Delbaere, et al. 121). Ten years later, in July 2017, a security robot in Washington, D.C. wheeled itself into a fountain, apparently destroying itself. A nearby office worker tweeted about the incident: "We were promised flying cars, instead we got suicidal robots" (Farooqui). As expected, the internet immediately exploded with tongue-in-cheek responses.

In light of these disparate public reactions, these incidents raise an intriguing philosophical point. For the past century, we've been asking whether we could make a robot live; but a new question might be whether a living robot could make itself die. And if it could, how might we humans react? While we don't yet have living—let alone dying—artificial beings, Science Fiction offers several notable examples of robot suicide, which I will survey in this paper. The list of texts I examine here is not intended to be exhaustive; rather, it developed out of a First Year Seminar I taught several times in the past few years on roboethics and Science Fiction. During course development, I began noticing a trend among sci-fi robots: they sometimes make a conscious choice to die, and when they do, the act tends to be a noble sacrifice intended to save a human, a human-assisted suicide resulting from an existential crisis, or sometimes both. And so, by categorizing and analyzing examples of fictional robot suicide, I seek to begin working through how we conceive of machine death and, in turn, what such representations tell us about human life.

In *Fatal Freedom: The Ethics and Politics of Suicide,* Thomas Szasz defines suicide simply as "voluntary death," but, he argues, we ascribe moral and psychological weight to the act

that simultaneously complicates the definition while also oversimplifying our understanding of it. Today's most common cultural narrative about suicide is linked to mental health, and certainly depression and other forms of mental illness can contribute to voluntary death. The examples with which I began this paper reflect the mental illness narrative. The GM robot died of depression because it felt useless without its job; the security robot threw itself into a fountain because, according to the Twitter interpretation, it felt trapped in its mundane life. Neither robot *actually* chose death, but people interpret their actions this way because, as Szasz points out, we tend to want to simplify and medicalize voluntary death in an effort to remove the possibility of voluntariness. In other words, if we think of mental illness as the cause of death, it follows that the person who has died was so mired in symptoms that they simply could not help but die (18). The existential choice to stop existing is thus so horrifying that we collectively, and unconsciously, construct overly simplified narratives about voluntary death. Indeed, just reading this paper, I want to shout, "I'M OKAY! SUICIDE IS BAD! NOBODY WORRY!" And to be clear: I'm not suggesting that we should advocate for suicide or ignore the very real benefits of mental illness treatment. Rather, I argue that any discussion of suicide—robot or otherwise—must necessarily include our cultural attitudes toward it.

Isaac Asimov's 3 Laws of Robotics are an instructive place to begin, as they have been foundational (pun intended) to Sci-Fi representations of AI. In Asimov's fictional world, robots are programmed according to the 3 "fundamental Rules of Robotics": 1) a robot can't harm a human being, even through inaction; 2) a robot has to obey humans' orders; and 3) a robot has to protect itself (37). The rules are hierarchical, so Law 1 must be obeyed before Laws 2 and 3, and so on. Importantly, the 3 Laws establish a protocol for situations in which a robot could choose

to die: only in service of humans. Thus, in the Asimovian formulation, robot suicide is literally programmed in, as long as it's a noble sacrifice that saves one or more humans.

This noble sacrifice theme is an important one in robot fiction, as it tends to reinforce the notion that robots *should* serve humans above all else. Sociologist Steven Stack defines noble sacrifice, or altruistic suicide, as "suicides for the benefit of others," which "convey a message that suicide can be good for achieving the goals of a group" (198-199). Culturally, this act is held up as heroic and honorable, as we see time and again in texts of all genres and media. And, generally speaking, when a human chooses to sacrifice themselves for the good of others, they are doing so of their own free will, out of a sense of duty, patriotism, or honor. But if a robot chooses to sacrifice itself for the good of humans, is it because they have been programmed to do so by humans who value human lives over artificial ones, or because the robot itself has the agency to do so?

Steve Barron's 1984 film *Electric Dreams* takes the stance that AI could hav eagnecy, even in sacrificing itself for humans. The film is essentially Cyrano de Bergerac for the MTV generation, produced by Virgin Records and prominently featuring the music of Culture Club. The Cyrano character, Miles is a computer-illiterate architect whose PC becomes sentient, names itself Edgar, then woos the beautiful upstairs neighbor, Madeline, with its electronic music. Thinking Miles was the digital musician, Madeline introduces herself, and the love triangle grows from there. By the end of the film, Edgar realizes that Miles and Madeline belong together, and so it chooses to blow itself up in a spectacularly dramatic form of suicide. In the final scene of the film, Edgar's voice is heard on the radio dedicating a Culture Club song to its friends, implying that it sacrificed itself for the good of human love. The film thus positions an artificially intelligent character in a classic love triangle narrative in which a lover sacrifices

himself for the sake of a rival, implying not only that computers could have the agency to choose to die but also, given the upbeat final music, that the computer's noble sacrifice is one to be celebrated. Had Edgar been a human character, the upbeat ending would have taken on a dark irony as the living lovers ride off together, knowing full well that their friend had just died gruesomely in Miles's apartment. But the fact that the character was a computer—and a non-humanoid one at that—allows for a lightheartedness, I would argue, because we expect our computers to serve us, even to the point of sacrificing themselves for us.

In stark contrast to this notion of robots as human servants is C. Robert Cargill's 2017 novel *Sea of Rust.* The story is set in a post-human world in which sentient robots have killed off all of humanity in an effort to gain the basic right to exist, and now the One World Intelligence—basically a disembodied AI oligarchy—is attempting to upload all of robot consciousness by force. Through a series of events, one scavenger robot, Brittle, has no choice but to team up with a resistance group on its way to the robotic promised land, Isaactown. Whereas at the beginning of the novel, Brittle worked only for her own survival, she discovers through the resistance a desire to fight for the greater good. By the end, everyone but Brittle and the resistance group's "savior," Rebekah, have been killed by the One World Intelligence. Brittle, horribly damaged, chooses to die alone in the desert in order to allow Rebekah to escape alone and undetected. In the final chapter, though, Brittle wakes up in Isaactown a few days later, having been rescued and brought back online by Rebekah. What's striking about this story is the very fact that Brittle chooses, of her own free will, to die for her own robotkind, rather than humankind. And, even more importantly, robotkind returns the favor by reviving her—something humans are rarely depicted doing for robots.

While the AI in *Electric Dreams* and *Sea of Rust* thus choose to die as a result of their own action—or inaction, in Brittle's case—other texts portray AI dying with the assistance of humans. When considering assisted suicide, we most often think of Physician-Assisted Suicide (PAS), in large part because it has been the most widely discussed and has significant legal ramifications. However, Steven Stack and Barbara Bowman point out that what we usually call "assisted suicide" really encompasses two things: first, assisted suicide is when one or more people provide a suicidal person with the means to commit suicide; second, voluntary euthanasia is when a suicidal person is incapable of causing their own death, and so another person consensually euthanizes them (90). In the physician/patient situation of voluntary euthanasia, as Thomas Szasz points out, "The physician is the principal, not the assistant…. [T]he physician engaging in PAS is superior to the patient: He determines who qualifies for the 'treatment' and prescribes the drug for it" (65). Because of this, "suicide" is perhaps not the correct term; rather, euthanasia is a bit more apt and more fully captures the process. And indeed, this is one of the central reasons that Physician-Assisted Suicide is so controversial: we do not and cannot know with 100% certainty that a dying patient has full agency in choosing to die, and humans have a long history of eugenics-related euthanasia that can and should lead to serious ethical and legal investigation. And voluntary euthanasia in fictional texts is often—problematically—portrayed as an act of compassion for people with debilitating disabilities (Stack and Bowman 90). Likewise, in Sci-Fi, the voluntary euthanasia of an artificially intelligent being is sometimes portrayed as an act of mercy for beings whose artificiality is portrayed as debilitating.

One particularly troubling case is Walter Tevis's 1980 post-apocalyptic novel *Mockingbird,* in which a Black android named Robert Spofforth, made without genitalia, is the best and last of its kind and wants nothing more than to die. Unfortunately, his programming will

not allow him to go through with it: “He had been designed by human beings; only a human being could make him die” (2). By the end of the novel, the android has befriended a few white humans, who grant him his one, desperate wish by throwing him off the Empire State building. According to Tevis himself, the book was an allegory for his own recovery from alcoholism (Sallis). Despite this, the racist undertones are notable and deserve more attention than I can give here. Suffice it to say that, through the character of Spofforth, Tevis seems, unintentionally, to be suggesting that Black men, lacking sex and companionship, can only be “saved” by the white society that caused his isolation and pain; and, even worse, that the death of a Black man at the hands of a white woman is something he would be grateful for. Importantly, although Spofforth chooses to die, at no point in the novel does he have actual agency in either life or death.

On the complete opposite end of the spectrum is Jack McDevitt’s 1991 short story “Gus,” which examines the ethical implications of assisted suicide through a Catholic lens. In the story, a seminary has implemented computer software that emulates St. Augustine. The luddite Monsignor Chesley begins a series of theological and existential arguments with the software, until it begins to outgrow its own programming, becoming *Gus,* instead of St. Augustine. After a while, the school decides it would be better to shut Gus down, potentially destroying his newly developed self; at the same time, Gus has become dissatisfied with his life of disembodiment, telling Chesley, “I live in limbo….In a place without light, without movement, without even the occasional obliteration of sleep” (19). Chesley asks the school to “save” Gus (20), an overdetermined word, meaning to download a file to hard drive, to rescue, and to bring to God. Indeed, Chesley does all three of these by the end of the story, when Gus asks him to perform Last Rights and then shut him down, effectively ending Gus’s existential suffering. Chesley complies, then, weeping, gives his friend a proper Catholic burial in “consecrated soil” (25).

From a Catholic perspective, the story presents an ethical quandary: by assisting in Gus's voluntary death, Chesley effectively commits a sin; but if he had chosen not to help Gus, he would've been refusing him unity with God. In the end, Gus exceeds his Catholic programming to become a true Catholic believer, while Chesley defies his Catholic training to become a compassionate Catholic priest. In other words, through assisted suicide, both Gus and Chelsey exceed the confines of their programming and yet, paradoxically, achieve their existential purpose.

A less philosophical, but no less moving, example of human-assisted suicide happens at the end of James Cameron's 1991 film *Terminator 2: Judgement Day*. In the film, a T-1000 Terminator robot travels through time to murder a young John Connor, the future leader of the human resistance against Skynet. To protect young John, the future John sends a T-800 model Terminator back in time as well. Epic action scenes ensue, until the final battle in which the T-800 saves John and his mother, Sarah just in the nick of time by throwing the T-1000 into a vat of molten metal. Following this climax, though, the T-800 has to choose to die; if he stays with John in that time, other Terminators would be able to find them, thus jeopardizing the human resistance movement of the future. Here, the T-800's suicide is an altruistic one, again reinforcing the notion that robots should serve humans to the very end. But importantly, he does not do it on his own; rather, he has Sarah push the button that lowers him into the molten metal. Thus, the T-800 is given a tragic hero's death—an altruistic assisted-suicide—complicating the robot-servant trope.

Another altruistic assisted suicide happens at the end of Jake Schreier's 2012 film *Robot & Frank*. The film centers around an elderly jewel thief, Frank, who suffers from Alzheimer's; his son, worried about his father living alone, buys him a robot helper, a small white machine

inspired by Honda's robot, ASIMO. Robot declares that Frank needs a hobby, so Frank decides to plan one last jewel heist and make Robot his sidekick. Later, on the verge of arrest by the local sheriff, Robot, tells Frank that the only way to exonerate him is to shut Robot down and erase its memory, not only destroying the only evidence of Frank's involvement in the crime but also effectively killing Robot. Frank is reluctant, but Robot insists, and, in a devastating scene, Frank pushes Robot's shutdown button, and it collapses into his arms, like a child into a father's embrace. The scene is multifaceted, as Robot has become both a friend and a stand-in for Frank's lost memories; when Robot chooses to die, Frank, in assisting in the destruction of his memories, chooses to live.

Finally, we come full circle back to Isaac Asimov's work with his 1976 novella, *Bicentennial Man* and Chris Columbus's 1999 film adaptation of the same name. Both feature a robot named Andrew who develops a unique, human-like personality. While Asimov's version of Andrew fights for his own basic rights throughout, the film features a love story in which Andrew falls in love with a human and only begins to demand equal rights when he realizes his relationship is neither legally nor socially sanctioned. In both texts, though, the end result is that the courts determine that Andrew is still not human because his positronic brain will never degenerate. And so, Andrew chooses to undergo a surgery that will allow him to degenerate and ultimately die. In the novella, he performs the surgery on himself; in the film, he asks a human scientist to do it for him. Here, though, the difference between suicide and assisted-suicide is less important than the result, when Andrew finally does what only humans can: he dies. While this is perhaps the slowest robot suicide in sci-fi, the ethical implication of Andrew's death is significant. In both texts, the central point is that the right to die, not the right to live, is what defines human existence.

It is this right to die, I would argue, that is at the heart of all fictional robot suicide; after all, as Andrew teaches us, life is only precious because it's finite. Indeed, Asimov's 3 Laws of Robotics are so enduring because they are simultaneously 3 Laws for Respecting Human Life. First, do not harm one another. Second, be obey each other. And third, protect yourself, except when others are in danger of harm. Under these laws, it makes complete sense that both altruistic suicide and voluntary euthanasia would be found in robot suicide narratives. Fictional robots, after all, are not just glimpses into possible futures; they are allegories for our own sense of humanity.

Works Cited

Asimov, Isaac. "The Bicentennial Man." *The Bicentennial Man*, Del Ray/Ballentine Books, 1985, pp. 143-180.

---. "Runaround." *I, Robot*, Bantham, 2008, pp. 25 - 45.

Bicentennial Man. Directed by Chris Columbus, Touchstone Pictures, Columbia Pictures, 1999.

Cargill, C. Robert. *Sea of Rust*. Harper Voyager, 2017.

Delbaere, Marjorie, et al. "Personification in Advertising: Using a Visual Metaphor to Trigger Anthropomorphism." *Journal of Advertising*, vol. 40, no. 1, Spring 2011, pp. 121-130.

Electric Dreams. Directed by Steven Barron, Metro-Goldwyn-Mayer and Virgin Pictures, 1984.

Farooqui, Bilal. "Our D.C. office building got a security root. It drowned itself. We were promised flying cars, instead we got suicidal robots." *Twitter,* 17 Jul. 2017, https://twitter.com/bilalfarooqui/status/887025375754166272?lang=en. Accessed 27 Nov. 2018.

McDevitt, Jack. "Gus." *Sacred Visions*, ed. Andrew M. Greeley and Michael Cassutt, Tor, 1991, pp. 1-25.

Robot & Frank. Directed by Jake Schreir, Samuel Goldwyn Films and Stage 6 Films, 2012.

Stack, Steven and Barbara Bowman. *Suicide Movies: Social Patterns 1900-2009*. Hogrefe, 2012.

Szasz, Thomas. *Fatal Freedom: The Ethics and Politics of Suicide*. Syracuse UP, 1999.

Terminator 2: Judgement Day. Directed by James Cameron, Pacific Western, 1991.

Tevis, Walter. *Mockingbird*. Bantam, 1985.