Android Dreams and Human Imaginings
by Aaron Barlow

What is not human? Does the fact of biological humanity divide “us” from everything else? What about something that acts with all of the intelligence, compassion, and sense of shared mission of the best human beings? What if these qualities were constructed, not born? Is such a being not human? What about someone who has withdrawn from knowledge, feelings, and any sort of interaction with others? Is that person no longer human? While these are good questions, they are not the only ones that science-fiction writers address when examining human relations with their creations; they are merely starting points. Such examinations lead back to a fundamental question: “What are we?”

Science-fiction writers, such as Alfred Bester, Philip K. Dick, Stanislaw Lem, and William Gibson, use robots, androids, and artificial worlds not to explore what is “out there” in fictional universes but to seek what is “in here,” in each of us. As Patricia Warrick writes, “Finding an answer to the question of what is truly human and what only masquerades as human is, for Dick, the most important difficulty facing us” (189). The same could be said of the other writers as well. Dick has written that nonhuman “creatures are among us, although morphologically they do not differ from us; we must not posit a difference of essence, but a difference of behavior. . . . A human being without the proper empathy or feeling is the same as an android built so as to lack it” (“Man, Android and Machine,” 202). As Joseph Francavilla puts it, “not only will the contrasts and oppositions between the human and the android become blurred, but . . . also the contrasting characteristics of each life form will switch sides. The human will become more inanimate; the nonhuman, more animate” (9). Understanding the implications and veracity of such a notion became these writers’ core endeavor.

Early on, science fiction in literature and film tended to interpret the relationship between humans and their creations in a paradigm of threat and opposition (such a paradigm stemming from golem tales and, in part, Mary Shelley’s 1818 novel Frankenstein, or The Modern Prometheus). Later, within a controlled framework of creation and mastery, the genre viewed human creations as something more “scientific” but still maintained an underlying dichotomy, difference, and subservient position within a hierarchy of control. Toward the beginning of the 1956 film Forbidden Planet, a crew, newly arrived on a distant planet, encounters a robot whose owner explains:

- Dr. Morbius. Don’t attribute feeling to him, gentlemen. Robby is simply a tool. Tremendously strong, of course. He could quite easily topple this house off its foundation.
- Doc. Morbius. No, Doctor, not even though I were the mad scientist of the tape thrillers because, you see, there happens to be a built-in safety factor. Commander, may I borrow that formidable-looking sidearm of yours? Thank you. Robby, point this thing at that Althaea frutex out there on the terrace. Fire. You understand the mechanism?
- Dr. Morbius. All right. Now turn around here. Point it at the commander. Aim right between the eyes. Fire. You see, he’s helpless, locked in a sub-electronic dilemma between my direct orders and his basic inhibitions against harming rational beings. Canceled. If I were to allow that to continue he would blow every circuit in his body.

Here, the creation (though not one made by humans, it turns out) is a mere extension of human will, its vast power completely in human hands. The dangerous creation has been brought under control, or, rather, the illusion of control.

The bringing to life of the inanimate has been a prominent theme for eons, the sculptor of antiquity Pygmalion being perhaps the most well-known classical example; the potential dangers, however, are more recently recognized. Not surprisingly, the most influential nineteenth-century discussion of the risks of such creation is Shelley’s Frankenstein. A little over a century later, Karel Capek’s play R.U.R. (1920)—the initials standing for “Rossum’s Universal Robots”—introduced the word “robot” and provided another examination of the creation’s possible dangers to its human engineers. Such tales, as literary scholar Istvan Csicsery-Ronay Jr. argues in a discussion of Ridley Scott’s 1982 film Blade Runner, show the slave either attaining equality with the Handy Man, refusing to consent to his domination, or revealing the illusory nature of the master-slave relationship. This shift entails the alien or subject creature asserting its equality via technological power. . . . All these deformations are condensed in Blade Runner’s Nexus 6. These Replicants were originally constructed to serve human purposes. As programmed beings, they had almost no free will; nonetheless, some were required to improve their functions as soldiers and sex slaves. They have sufficient freedom to become aware of their predicament. (253)
The difference is that the Replicants are no longer the enemy, except when humans make them so, and they have increased their free will, becoming as human as many humans.

In 1939, a tale called “I, Robot”—attributed to Eando Binder, the pen name of the brothers Earl and Otto Binder—appeared in Amazing Stories magazine. The story focuses on a self-aware robot called Adam Link, and attempts to balance earlier conceptions of the hazards of human creation by presenting a case in which a robot is wrongly perceived as evil. Unjustly accused of killing its creator, the robot eventually decides that it is not worth trying to clear itself and prepares to deactivate. Soon after this story appeared, science-fiction writer Isaac Asimov began a robot-themed series for Astonishing Stories, Super Science Stories, and, most importantly, Astounding Science Fiction. Asimov believed that any “real” robots would need to be designed so that they could not threaten humans.

In these stories, Asimov develops what came to be known as “the three laws of robotics,” which appeared for the first time in his 1942 story “Runaround” in Astounding Science Fiction. Two characters discuss these rules:

“Now, look, let’s start with the fundamental Rules of Robotics—the three rules that are built most deeply into a robot’s positronic brain.” In the darkness, his gloved fingers ticked off each point.
“We have: One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm.”
“Right!”
“Two,” continued Powell, “a robot must obey the orders given it by human beings except where such orders would conflict with the First Law.”
“Right!”
“And three, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.”
“Right! Now where are we?” (36–37)

We have been asking that same question in relation to robots, androids, and other creations (real and imagined) ever since. Though Robby of Forbidden Planet—whose name is taken from one of Asimov’s stories—follows the three laws, we have realized since his appearance that the laws are based on a simplistic hierarchy of human and machine that is unlikely ever to exist, a relationship that Csicsery-Ronay Jr. sees as that of “the Willing Slave.” He writes, “The selfhood of the Willing Slave is, of course, a subset of his Master’s, and will not extend beyond that the boundaries that the [master] inscribes. But he is independently conscious to the degree that he can learn higher rationality, and willingly choose his subaltern position” (Seven Beauties of Fiction, 229). As robots, the Willing Slaves “are aware that it is possible to act against the interests of human beings. Yet they cannot” (252).

French writer and filmmaker Emmanuel Carrère claims:

From the earliest science fiction on, the robot—like the golem and Frankenstein’s monster before it—had been cast in the role of villain, its human creator’s most cunning adversary. . . . Isaac Asimov had tried to impose a code of good conduct on robots and their writer-creators, to reduce the theme of robot rebellion to the scientific absurdity and cheap literary convention it was, but he did not succeed. (131)

Just as Asimov, the Binders, and other early science-fiction writers found faults in the traditions of Frankenstein, other writers soon began to chaff under the simplicity of the three laws and worry about the conundrum of androids or robots being both able and unable to act against humans.

Francavilla sees androids as “constructions of organic artificial life resembling humans” (as in the Replicants of Blade Runner) and robots as “inorganic artificial life resembling humans which is fundamentally a machine or mechanical being” (7). In either case, the resemblance of robots to humans raises another question, one not just of physical resemblance but of mind. This question was brought to the attention of science-fiction writers through mathematician Alan Turing’s 1950 essay “Computing Machinery and Intelligence.” Turing begins: “I propose to consider the question, ‘Can machines think?’” (433). This question piqued the interest of the science-fiction community, especially. Carrère writes:

Turing takes up the range of objections that had been raised against the possibility of artificial intelligence—that what computers do is too specialized to be called thinking, that they lack spontaneity, moral sense, desire, and taste, and so forth. Turing dispatches these arguments one by one, and proposes instead a single criterion by which to answer the question of whether a machine can think. That criterion is whether the machine is capable of making a human being believe that it thinks as he does. (132)

Not only did Turing provide new legitimacy to an area that science-fiction writers were exploring, but his work also led to the so-called Turing test, in which a human is presented with two responses to one question. If that person cannot tell which response comes from a machine and which comes from another human, the machine has passed the test.
In response to Turing, science-fiction writers began to wonder about not just how to control androids and robots, who were sure to eventually “think,” but also how to understand the community that humans and their creations were sure to become one day. Among these writers was Alfred Bester, whose “Fondly Fahrenheit” appeared in *The Magazine of Fantasy and of Science Fiction* in 1954. In the story, the relationship between human and machine becomes much more complex than anything found in Asimov. The last lines are particularly instructive:

Vandaleur didn’t die. I got away. They missed him while they watched the android caper and die. But I don’t know which of us he is these days. Projection, Wanda warned me. Projection, Nan Webb told him. If you live with a crazy machine long enough, I become crazy too. Reet!

But we know the truth. We know that they were wrong. The new robot and Vandaleur know that because the new robot’s started twitching too. Reet! Here on cold Pollux, the robot is twitching and singing. No heat, but my fingers write. No heat, but it’s taken the little Talley girl off for a solitary walk. A cheap labor robot. A servo-mechanism . . . all I could afford . . . but it’s twitching and humming and walking alone with the child somewhere and I can’t find them. Christ! Vandaleur can’t find me before it’s too late. Cool and discreet, honey, in the dancing frost while the thermometer registers 10° fondly Fahrenheit. (489)

With its structure—the narrator moving from human to robot to both, seemingly at random—Bester’s story questions the possibility of an absolute distinction between owner and the property owned, between humanity and those it has created to “serve.” Such ideas set the stage for the next two decades and the complex explorations of such writers as Dick, Lem, and Gibson. The distinctions and lack thereof between humanity and machinery have had significant impacts on science-fiction film, including works as divergent as Stanley Kubrick’s *2001: A Space Odyssey* (1968), Ridley Scott’s *Blade Runner*, Steven Lisberger’s *Tron* (1982) and Joseph Kosinski’s *Tron: Legacy* (2010), Mamoru Oshii’s *Ghost in the Shell* (1995), and Andy and Larry Wachowski’s *Matrix* trilogy (1999; 2003; 2003).

Perhaps the most influential science-fiction writer in terms of attitudes toward human and machine relations, Dick still had ambivalent feelings with regard to the relations between people and their creations. This ambivalence is evident in his earliest stories, including "The Defenders," which appeared in *Galaxy* magazine in January 1953 and was later incorporated into Dick’s novel *The Penultimate Truth* (1964). A war has driven both Americans and Russians underground, leaving the conduct of the war to “leadies,” robots able to survive the highly radiated surface. One of them speaks to a group of humans it is trying to block from returning to the surface: "I am sorry," the leader said, “but it is for your own protection. We are watching over you, literally. You must stay below and let us conduct the war. In a sense, it has come to be *our* war. We must fight it as we see fit" (78). The humans manage to turn the tables on the leadies and get to the surface, which they find is not radioactive but has been restored as a pristine human environment:

“*But why?*” Taylor asked, dazed. *He stared down at the vast valley below. “Why?*”

“You created us,” the leady said, “to pursue the war for you, while you human beings went below the ground in order to survive. But before we could continue the war, it was necessary to analyze it to determine what its purpose was. We did this, and we found that it had no purpose, except, perhaps, in terms of human needs. Even this was questionable.” (80)

Created with one vision and purpose, created with a specific desire built-in, the leadies reimagined that desire—even including human desire alongside their own—for the sake of the humans.

In “The Defenders,” Dick puts his finger directly on the great weakness of Asimov’s three laws: robot “desires” are not necessarily the same as those of their creators. No matter the restrictions placed on them, robots will work to fulfill their own desires. But how, one may ask, can a robot desire? It lacks the self-awareness of humans, and therefore lacks feelings. To Turing, however, this question may be irrelevant. If a robot seems to desire, it desires. Furthermore, desires are not simply feelings or caught up with volition and will—capacities that we associate with human beings and not with machines. Other entities certainly desire, as plants desire the sun, turning their flowers to catch its rays. Desire encompasses much more than human yearning and, as with the leadies, can be built into machines: a streetcar desires to run along its tracks. If it does not, it is no longer a streetcar.

In another of his early stories, "The Short Happy Life of the Brown Oxford," Dick approaches this issue from another angle. Instead of portraying desire as crucial to life or the power of decision, he imagines desire as an irritation. The character Doc Labyrinth explains:

“The Principle came to me this way. One day I was sitting on a rock at the beach. The sun was shining and it was very hot. I was perspiring and quite uncomfortable. All at once a pebble next to me got up and crawled off. The heat of the sun had annoyed it."

“Really? A pebble?”

“At once the realization of the Principle of Sufficient Irritation came to me. Here was the origin of life. Eons ago, in the remote past, a bit of inanimate matter had become so irritated by something that it crawled...
away, moved by indignation. Here was my life work: to discover the perfect irritant, annoying enough to bring inanimate matter to life." (250)

Humor aside, the anthropomorphism that Dick lampoons (and that Turing sidesteps) can be a real problem in coming to an understanding of human and machine relations—but that does not mean that such human traits as desire do not have machine analogues.

The leads of “The Defenders” desire to do the best for the humans they are programmed to serve, even if that means fooling the humans into circumventing their warlike nature to establish peace. Whatever the ontology of their desires, the leads end up as independent actors, making their own decisions and following plans of their own design. They act as though they have will of their own. Indeed, they would pass any Turing test that attempted to differentiate their responses from human ones.

Turing’s logic and Dick’s stories lead one to believe that robots, androids, and other creations in fact could become independent actors with individual volition. Would nothing make them different from human beings? Would they have individual consciousness behind that will? Would they desire in the ways that humans desire, or would their desires remain a mere metaphor of ours? To Turing, such questions probably would have no relevance; to science-fiction writers like Dick and Lem, they are fascinating.

With the exception of “The Short Happy Life of the Brown Oxford,” Lem generally has more fun with human and machine relations than Dick. Part of Lem’s intent is to lampoon the clichés of the science-fiction genre and beyond (similarly to how English writer Douglas Adams would create his own satire one generation later in his Hitchhiker’s Guide series), but he also points to the serious consequences of ill-conceived human and machine interactions. His characters are often powerful, whereas Dick’s characters generally are people oppressed in some way. The “constructors” of Lem’s Cyberiad (1974), Trurl and Klapaucius, succeed; Dick’s characters reach but rarely manage to grasp.

A focus on the nuances of human and machine relations is quite clear in Lem’s stories featuring Pirx, a spaceship pilot; these stories are collected as Tales of Pirx the Pilot (1979) and More Tales of Pirx the Pilot (1982). As literary historian Jerzy Jarzębski submits:

inquiring in the person of Pirx Lem intended to test mankind. He set out to find a place, in the world of triumphant technology, where human weakness and human imperfection are no longer defects. It is simple to say: “Machines cannot think, machines have no consciousness!” From where do we derive this certainty? . . . Whence come the criteria that allow us to define “thought” in an apodictic way—and what if not thinking shall we call machine operations that have analogous results for their “output”? Lem’s robots are not primitive machines; they represent a serious challenge to man. (120)

Ultimately, for both Dick and Lem, the questions should be turned around, away from the robots and back to human beings. What makes us human? What makes us fundamentally different from self-aware machines? Over the course of the half-century since Bester’s story, the attempts to answer such questions have provided some of the most intriguing work of the science-fiction genre.

For Dick, Lem, and Gibson, the questions become even more complex, for the very natures of humans and machines are dependent on environment. How can one be truly human in an inhuman environment? How can one be real in an unreal world? In I Hope I Shall Arrive Soon (1985), Dick writes:

I consider that the matter of defining what is real—that is a serious topic, even a vital topic. And in there somewhere is the other topic, the definition of the authentic human. Because the bombardment of pseudo-realities begins to produce inauthentic humans very quickly, spurious humans—as fake as the data pressing at them from all sides. My two topics are really one topic; they unite at this point. Fake realities will create fake humans. Or fake humans will generate fake realities and then sell them to other humans, turning them, eventually, into forgeries of themselves. So we wind up with fake humans inventing fake realities and then peddling them to other fake humans. (6)

If a human can become a fake human, why cannot a fake human, an android, become a real human? Where, in other words, does humanity lie? This is something Turing does not bother to answer. Is it simply something physical, or is there also a necessary spiritual aspect?

Perhaps these questions miss the point. Toward the end of Do Androids Dream of Electric Sheep? (1968), Rick Deckard remarks, “The electric things have their lives, too. Paltry as those lives are” (241). With this statement, Deckard erases the distinction between human and machine, and with it the notion that humanity constitutes an exclusive and discrete set of qualities. Yet Dick cannot let the question go and continues to explore it throughout his career—even though he clearly knows that, on some level, Turing and Deckard are right. In the 1980s, Gibson took a different approach, exploring the interface of humanity and machines in Neuromancer (1984), Count Zero (1987), and Mona Lisa Overdrive (1988), novels in which the two meld in a “cyberspace” that is separate from “normal” reality—something of a sensorial Internet accessed
through computer portals. For Gibson, the intersection of machine life and human life allows both to bleed into each other. Robert Longo’s 1995 film *Johnny Mnemonic* reflects this blurred relationship, as information stored in the title character’s brain threatens his memories and his life.

Of course, Gibson’s most important contribution to discussions of humanity and machines comes from his concept of cyberspace, also known as virtual reality. In his article “The Sentimental Futurist: Cybernetics and Art in William Gibson’s *Neuromancer*,” Istvan Csicsery-Ronay Jr. sees cyberspace primarily in terms of science fiction. He claims that by developing the concept

Gibson restored the heroic spatial expanse that SF had lost in outer space and laid the groundwork for developing a system of symbols for cybernetic implosion. Cyberspace, after all, is a purely human system, a “consensual hallucination,” with no objective status. . . . It cannot be conquered for humanity because it is an aspect of humanity at the outset. The best that can be hoped for is its conquest for unalienated, enlightened human beings from the powers of avarice, fetishization, and global reification that control the cyberspace field. (223–24)

Though humanity has yet to conquer cyberspace, things have changed since Csicsery-Ronay Jr.’s article. The Internet has achieved a position of objective status. The “consensual hallucination” has become a human reality, but it is one that we still do not completely understand. Through its social networks and its tendency to isolate individuals behind screens, the impact of this reality on our sense of humanity remains to be seen. Whatever the case may be, the Internet is certainly changing ideas of our relations to machines, as we become more and more dependent on the mechanical aspect of our lives, which already anticipates the efficiency of Dick’s leadies.

There may be another Turing test yet to be developed, one that does not test human intelligence but rather human empathy—a test that addresses Dick’s concern for humans drifting away from humanity, not machines’ movement toward it. In one of his last novels, *The Divine Invasion* (1981), Dick writes of the “Beside-Helper,” who offers to assist the newly dead with the test of judgment to come:

He offered to present his own bill of particulars to the retribution mechanism in place of the bill of particulars of the person. If the person were innocent this would make no difference but, for the guilty, it would yield up a sentence of exculpation rather than guilt. . . .

“What does the Beside-Helper’s bill of particulars list?” he asked.

“It is blank.” . . .

“The retributive machinery could not process that.” . . .

“It would process it. It would imagine that it had received a compilation of a totally spotless person.” (127–28)

Everybody believes, however, that they are more good than bad, so the balance is already in their favor.

The way in which Dick sets up this situation—with “retributive machinery” making the decision and not a human—turns the Turing test on its head, much like CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) tests, and proves that online responses are supplied by humans, not machines. In *Blade Runner*, Scott conveys an understanding of this turnaround, and incorporates it into the film’s ending, when Deckard (a human who may be a Replicant) and Rachel (a Replicant who has been led to believe she is human) exit together. Literary scholar Judith Kerman writes:

I do not hear the closing elevator doors that end the Director’s Cut of *Blade Runner* as the clap of doom for Deckard and Rachel. . . . They have been expelled even from the dubious Eden of protection by the forces of Tyrell and his police apparatus, and now they must earn their bread in a hard world by the sweat of their brows. As in our own society and our own lives, we do not know the end of *Blade Runner*, but we can anticipate that the hard road of . . . reality lies ahead. (39)

Yet Deckard and Rachel seem to be among the few who have taken on the assistance of the Beside-Helper: They are going into their future with a clean slate, the distinctions of human and machine erased.

This is a far cry from the end of Shelley’s *Frankenstein*, which includes the last words of Victor Frankenstein’s abandoned creation:

“Forwell, Frankenstein! If thou wert yet alive and yet cherished a desire of revenge against me, it would be better satiated in my life than in my destruction. But it was not so; thou didst seek my extinction, that I might not cause greater wretchedness. . . . Blasted as thou wert, my agony was still superior to thine; for the bitter sting of remorse will not cease to rankle in my wounds until death shall close them for ever.
But soon," he cried, with sad and solemn enthusiasm, "I shall die, and what I now feel be no longer felt. Soon these burning miseries will be extinct. . . . My spirit will sleep in peace; or if it thinks, it will not surely think thus. Farewell."

He sprung from the cabin-window, as he said this, upon the ice-raft which lay close to the vessel. He was soon borne away by the waves, and lost in darkness and distance. (225)

Blade Runner's parallel comes earlier, in the death of the android Roy Baty, the inheritor of the mantle of Frankenstein monster's. Similarly aware of his own being and the circumstances of his creation, Baty is torn between the strength of his being and the weakness of his origin. In Blade Runner, however, the road does not end at the point of the creation's death but continues. While Deckard defeats his own "monster," something different from any past relationship between human and machine is portended as he and Rachel exit.

As it winds into unknown territory, Deckard and Rachel's path also winds back to ancient human mythology. Ovid's Metamorphoses, written over two thousand years ago, contains the tale of Pygmalion, who weds an ivory statue of his own creation after it comes to life. Perhaps, in our contemporary vision of human and machine relationships, the same good life might meet Deckard and Rachel, and the audience as well.

Works Cited

_____.


