Hey, Robot!

Marc Donner
Associate Editor in Chief

What area of research, development, and commercial activity owes more of its existence to the arts than robotics does? The word itself comes from an early 20th-century play; less than a decade later, an important film introduced an enduring fantasy concept of what robots look like. Shortly after that, but still before much significant technical research or development occurred in the field, science-fiction writers developed complex theories of robot behavior in stories that are still in print today.

In this installment of Biblio Tech, we'll look at some of the arts that have shaped our notions of robots. We will see the deep roots these stories have in far earlier concepts that have little to do with engineering but everything to do with the human race's fascination with creation.

R.U.R. (Rossum's Universal Robots)

In 1920, Karel Capek completed his play, *R.U.R. (Rossum's Universal Robots)*; its first production in 1921 brought Capek worldwide renown and introduced the word "robot" to the English language. Some argue that if he had survived the era of Nazi domination in Europe, he would have received the Nobel Prize for Literature.

Rather than mechanical constructions, Rossum's robots were more biological and chemical in their fabrication. Nonetheless, they are definitely the ancestors of our modern industrial gadgets. Is the distinction between human beings and machines that humans work to live while machines exist to work? If that's the case, then Rossum's robots definitely existed, or were at least built, to work. They worked tirelessly and were tremendously more productive than mere humans, but they lacked emotions, creativity, and souls.

In the play's first act, Helena Glory, the young daughter of "the President," arrives by ship to the remote island where Rossum has developed the techniques for making robots. She is concerned with the oppression of robots worldwide and wants to foment a revolt among them—to inspire in them a passion for freedom.

What she finds on the island is a factory almost entirely staffed by robots, with a small team of men managing the operation. She is dismayed to discover that the robots are emotionless and unmovable: Rossum and son's original engineering work produced a simplified physiology and nervous system that were incapable of pain or passion.

All is not lost, however. Dr. Gall, the head of the Experimental Department, is working to add a pain sense to the robots:

> Helena: Why do you want to cause them pain?
>
> Dr. Gall: For industrial reasons, Miss Glory. Sometimes a Robot does damage to himself because it doesn't hurt him. He puts his hand into the machine, breaks his finger, smashes his head, it's all the same to him. We must provide them with pain. That's an automatic protection against damage.

In addition, there's a mysterious disease called "Robot's cramp" that the managers view as a fatal failure: "A flaw in the works that has to be removed." Helena recognizes it as something else, though: "No, no, that's the soul."

In the remainder of the play, we watch the world's economies devastated by cheap labor and see governments wage war with armies of robot soldiers. Finally, the robots revolt, ultimately exterminating their creators. The play ends with the emergence of a robot Adam and Eve and the cycle of life begins again.

Frankenstein, the Golem, *and* Metropolis

We see in Mary Shelley's 1818 novel *Frankenstein* (unlike the flood of cinematic caricatures that sprang from it) a set of concepts similar to those in *R.U.R.* Behind Frankenstein, we see the even older legend of the Golem. In 16th-century Prague, the story goes, Rabbi Loew created a humanoid figure out of clay and brought it to life by marking it with a powerful magic word. He then commanded this creature to defend the Jews of the Prague ghetto against the torments of a contemporary despot. Ultimately,

What is a robot?

There is no real consensus on precisely what a robot is. Rather than trying to define one, let's instead try to identify the characteristics of things that we might call robots. The most appealing description is that a robot is a system with mechanical components intended to achieve physical action; it also has sensory feedback and a sophisticated and flexible control system that links its sensing to action.

To see if this characterization works, let's see if it correctly distinguishes between our ideas of robots and nonrobots. The system must be intended to produce mechanical action, so a computer video game is out. The system must use sensory feedback to control motion, so printers are out. So far, there is little to distinguish a robot from a classical control system.

A numerically controlled machine tool is a robot, but just barely. A modern car's antilock braking system could qualify, although there's something unsatisfying in it doing so. An airplane's autopilot certainly qualifies as a robot, particularly the advanced autopilots that can receive a list of waypoints and then navigate themselves from liftoff to approach via GPS. A washing machine that can sense the amount and temperature of water in its tub and act accordingly is probably a robot, albeit not a particularly interesting one. Oddly enough, many of the pick-and-place industrial robot systems in factories in Japan and elsewhere around the world fail this test, because they lack a sensory capability.

Today, robots are almost commonplace. We see them in numerous prosaic roles in factories, but we also see them competing in what can best be called a new form of demolition derby. Students around the world work to build robots to compete in robot soccer; a researcher at Bell Labs built one to play ping-pong a few years back. Numerous toys on the market incorporate various aspects of robotic technology.

the creature began to show signs of rebellion. The ending of the legend has many different variations. In some versions, Rabbi Loew destroys the Golem; in others, the creature flees, never to be seen again.

Fritz Lang's 1927 film *Metropolis* introduced the first cinematic robot, which managed to typecast the entire category for at least 50 years. Lang's robot is the creation of a mad scientist, Rotwang, who is trying to create a surrogate for his lost love, Hel, to whom he has built an altar in his laboratory. She rejected him in favor of his rival, Joh Fredersen, and died giving birth to their son, Freder Fredersen. Joh Fredersen is the master of the city of Metropolis, an architectural and industrial vision of the early 20th century that we might barely recognize today. Metropolis is divided into two parts: a lower part inhabited by industrial workers who live underground and toil ceaselessly in the bowels of the machines that make Metropolis function, and an upper part peopled by a happy leisure class who spend their time at games and diversions. Near the beginning of the film, Freder ventures below ground, where his heart is moved by the plight of the workers and captivated by the beautiful Maria, a pure and gentle young woman whom he encounters preaching peaceful change. She promises a bridge for the gap between the workers (the Hands) and the managers (the Head). She calls this as-yet-unknown person the Mediator and identifies him as the Heart.

To undermine the workers' movement, Joh has Rotwang give the robot Maria's appearance. The robot then proceeds to rouse the workers to violence, which backfires when their children are threatened by floods unleashed by the destruction of some of the machines. Freder and the real Maria rescue the children, and the mob then burns the robot at the stake as Freder brokers a reconciliation between Joh and their leader.

The robot is referred to as the Machine–Man in the English intertitles before it is transformed into Maria's sinister double. The double is everything that a thousand subsequent movie robots ever were: destructive, soulless, and ultimately evil. This movie is one of the most influential achievements of 20th-century filmmaking; you can see its influences in many subsequent cinematic masterpieces, as well as nearly every third-rate monster flick.

A common theme running through all these early stories is the classical Promethean notion that certain things are not meant for humans to control. Tampering with them trespasses on the domain of the divine and exposes the trespasser to severe punishment. Mary Shelley, in the preface to the 1831 edition of *Frankenstein*, wrote,

> "Frightful must it be; for supremely frightful would be the effect of any human endeavor to mock the stupendous mechanism of the Creator of the world."

Why is it that these stories—from the Golem to *Frankenstein* to *Metropolis*—always adopt classical models? Creating something that is alive or seems to be alive is portrayed always as trespassing on the perquisites of the divine, which is hubris and is certain to be punished by the gods. A simple explanation is that every storyteller tries to create a fiction that meshes with the real world—in this case, a real world in which intelligent robots are manifestly absent. To be complete, then, each story must end with a world without such things and a reason for their absence. You might ask, "But why aren't there any manmade intelligences?" to which the answer would be, "Because there shouldn't be, of course."

In the 20th century, however, technological progress started to undermine the tyranny of "cannot." Let's look at the effect of that change on "should not."

I, Robot

In 1939, a young man with a BS in chemistry from Columbia University wrote a story called "Robbie" about a little girl's robot playmate. In a retrospective article about this and his other robot stories, entitled "My Robots," Isaac Asimov said,

> "In that case, what did I make my robots? I made them engineering devices. I made them tools. I made them machines to serve human ends. And I made them objects with built-in safety features. In other words, I set it up so that a robot could not kill his creator, and having outlawed that heavily overused plot, I was free to consider other, more rational consequences."

In making them "to serve human ends," Asimov didn't innovate. However, in delving more deeply into their construction, particularly into their cognitive construction, he broke new ground.

Asimov went on to earn a PhD in biochemistry and work in academia teaching science, all the while writing science fiction throughout his long career. He brought to his writing tremendous conceptual power and a deep theoretical orientation. He was renowned as a prolific writer who could turn out a story or a book in a startlingly short time period, but this speed came at the expense of quality.

Much of Asimov's early writing was not his best. The characters in the short stories that make up *I, Robot* are flat, the dialogue wooden, and the best of the plots contrived. He did have exceptional moments in those early days when his writing soared—for example, in *Nightfall*—but in his youthful work this was the exception rather than the rule.

Nevertheless, the stories in *I, Robot* are important works, because in addition to repudiating the divine "You may not mock the stupendous mechanism of the Creator of the world" taboo, Asimov made a more fundamental contribution—namely, the Laws of Robotics:

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 to 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 Laws.

Suddenly, the issue was not about the sin of creating robots: it was about how to manage them appropriately. The importance of Asimov's Laws of Robotics was not their precise formulation or wording, but that they existed at all. Engineered things, Asimov tells us, can be made subject to strict controls that aren't applicable to humans. This is why constructing robot intelligence is not a sin, he says, any more than constructing anything else is a sin. Yes, we must pay attention to complicated details, but difficulty is not impossibility. Check with the Wright brothers, Sir Edmund Hillary, and one or two others if you doubt that fact.

Bolo

With *Bolo*, Keith Laumer introduced the robot's viewpoint. In this series of stories begun in 1960 and largely complete by 1969, we encounter a series of robotic war machines—the evolutionary descendants of tanks. Laumer wasn't the theoretician that Asimov was, and the logic driving his thinking isn't particularly transparent, but the concept is compelling.

In the *Bolo* stories, Laumer inserts sections of first-person monologue by the robot. This is a big step away from Lang's notion of the robot as incomprehensibly alien—the Other. Instead, the robot thinks about its sit-

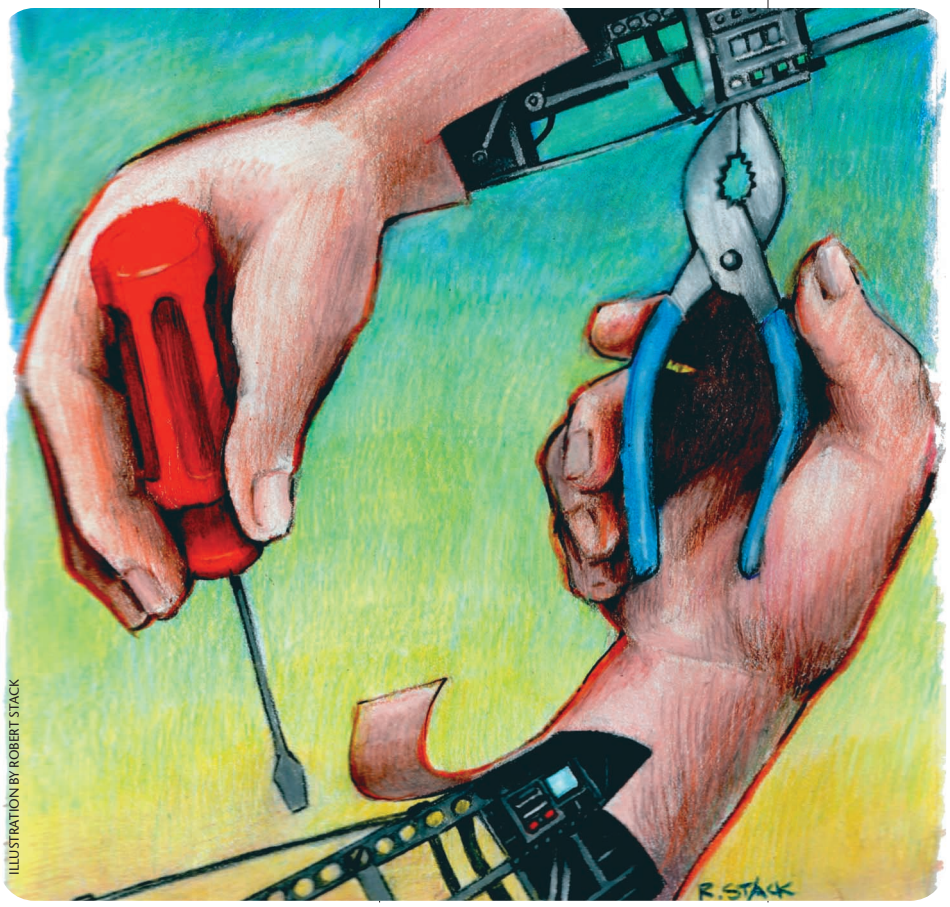

ILLUSTRATION BY ROBERT STACK

Influential Works

MEDIUM	TITLE	AUTHOR	ORIGINAL APPEARANCE
Book	*Frankenstein*	Mary Shelley	1818
Play	*R.U.R. (Rossum's Universal Robots)*	Karel Capek	1920
Film	*Der Golem* (in German)	Paul Wegener, Carl Boese	1920
Film	*Metropolis*	Fritz Lang	1927
Book	*I, Robot*	Isaac Asimov	Short stories, 1940–1950; collection, 1950
Book	*Nightfall*	Isaac Asimov	1941
Book	*Bolo*	Keith Laumer	Short stories, 1960–1976; various collections
Book	*Do Androids Dream of Electric Sheep?*	Philip K. Dick	1968
Film	*Star Wars*	George Lucas	1977
Film	*Blade Runner*	Ridley Scott	1982
Book	*The Amazing Adventures of Kavalier and Clay*	Michael Chabon	2000

uation and reasons about the circumstances in which it finds itself. Laumer's robots are invariably loyal to their human masters, although in "Rogue Bolo," we encounter a robot with sufficient intellectual power to conduct a strategic campaign against adversaries that humans haven't detected, despite direct orders from humans to desist. Implicit in this is Asimov's assertion of the First Law's precedence over the Second Law.

Star Wars

Released in 1977, *Star Wars* struck another small blow in the struggle to liberate robots from their earlier stereotypes as humanoid, ruthlessly competent, and evil. In this movie, we get a humanoid robot—C3PO—that is trivial and cowardly, though still part of the good guy crowd, in contrast to the lumpish and purely functional (but invariably competent and heroic) R2D2. C3PO is articulate whereas R2D2 is completely wordless, thus providing the ultimate cinematic example of the old saw that actions speak louder than words. Interestingly, in the recently released back-story, *The Phantom Menace*, we learn that the young Anakin Skywalker constructed C3PO. R2D2's origin seems to be more prosaic, but there is some sort of justice in the fact that the weak C3PO was built by the person who turns out to be the penultimate bad guy. Perhaps C3PO's weakness is a foreshadowing of Anakin's own? It's worth noting that by 1977, robots were so well established that this convergence of two separate themes—the fiction-inspired C3PO and the reality-inspired R2D2—merits no more than a miniscule side plot in a science-fiction film.

Blade Runner

With Ridley Scott's 1982 movie *Blade Runner*, loosely based on Philip K. Dick's 1968 novel *Do Androids Dream of Electric Sheep?*, we return from mechanical humanoids to the biological creations Capek pioneered in *R.U.R.* Superficially, this is an exercise in which androids, called replicants and physically indistinguishable from humans, rebel against a social order that treats them brutally. They are banned from Earth—a formula that Asimov used to great effect in his robot stories—and have artificially limited lifespans. Their superhuman physical and mental capabilities are key to detecting them when they run and hide. We learn in *Blade Runner* that they fear death and so run to seek freedom and an unimpeded lifespan.

The understatement built into *Blade Runner* is overwhelming. Are replicants human? Their bodies are biological and they look like people, so it's too easy to grant them souls by dismissing their creation as some perversion of cloning. Hannibal Chew, the engineer who boasts, "I design your eyes," to two replicants right before they kill him, refutes this: if he'd just cloned their eyes, how could they have superior eyesight? And Harrison Ford's character, Deckard, is a paradox: How can a human, every other instance of whom is manifestly inferior to replicants in physical and intellectual capabilities, manage—unaided—to defeat an entire team of replicants, one after another?

Moreover, *Blade Runner* re-poses the same question that *R.U.R.* asked: can you create an entity with intellectual capabilities and not give it a soul? Deckard speculates on this at the end of the movie while reflecting on a replicant's decision not to kill him when he'd won the final fight:

> "I don't know why he saved my life. Maybe in those last moments he loved life more than he ever had before. Not just his life, anybody's life, my life. All he'd wanted were the same answers the rest of us want. Where did I come from? Where am I going? How long have I got? All I could do was sit there and watch him die."

Meanwhile, in the real world…

The golden age of robotics research

came to an end sometime in the mid 1980s when a pair of economists observed that the sweet spot for flexible automation was in an area in which US industry took no interest. It turned out that robots are cost-effective for production runs roughly between 1,000 and 10,000 units. US manufacturing tends to have its sweet spots below 1,000 (airliners, electric generators, and supercomputers) and above 100,000 (jelly beans and automobiles). Japan's manufacturing industry has historically focused its attention on 1,000 to 10,000 unit runs, giving it a tremendous ability to respond to market dynamics with revised products and simultaneously making robotics a far more economically attractive proposition. The result of this economic insight was a dramatic drop in research funding for robotics worldwide. Nonetheless, the field has made substantial technical progress in the past 20 years, albeit largely out of the public eye. Interestingly, there hasn't been the same attention to robotics in the science fiction community, at least not in the works that have gotten attention from the broadest community of readers.

Is this parallel drop-off in the world of fictional robots because Asimov and Laumer said everything there is to say about robots? Is it that people have recognized the absurdity of humanoid robots and have transformed the debate into one about the broader topic of artificial intelligence, as we considered in the first Biblio Tech article? Or are we just bored with the topic? I'm not sure. I prefer to think that we're just waiting for some powerful new talent to turn our thinking on its head again with a brilliant new insight. □

Marc Donner *is an executive director in the Institutional Securities division of Morgan Stanley where he focuses on system and data architecture around client relationships. He is a member of the IEEE Computer Society and Usenix. Contact him at donner@tinho.net.*

Recruiting for *IEEE Security & Privacy* Magazine Conference Reporters

With so many security and privacy conferences these days, who can keep up with them all?

IEEE Security & Privacy magazine would like to help its readers by providing concise, informative summaries of significant events at conferences and workshops. You can help by volunteering to provide summaries of important papers, discussions, and events from the meetings you attend. Here are some guidelines, if you or someone you know is interested. You can also send queries to clandweh@nsf.gov.

- You are writing for the readers of a magazine. Think of yourself as a member of the audience—if you couldn't attend the meeting, what would you like to know about it from a friend who was there? Put the most important things first and be brief. Write in the active voice.
- Most readers are interested in significant technical advances. For most of the meetings *IEEE Security & Privacy* will cover, however, it isn't necessary to cover those details in depth because the IEEE Computer Society usually publishes a proceedings containing all the technical papers,. The questions asked after a paper and un-minuted panel discussions usually deserve more space than a rehearsal of the papers' abstracts. It is helpful to note what caught the audience's interest (or what didn't, if that's significant).
- It's helpful to let readers know how to acquire a copy of the proceedings—try to provide a reference; pointers to Web pages are good, if available.
- Details of the meeting outside the technical sessions can liven up the story. We aren't looking for gossip, but who won the croquet tournament could be of interest. Was the attendance up or down from last year? What are the plans for next year—dates, location, points of contact?
- It takes some work, but it can be rewarding to you as well as to readers. Reporting on a meeting as a whole forces you to look at it with a somewhat broader perspective than if you were just listening for the points that directly affect your own research.
- Try to get your copy in as soon as possible. "News" ceases to be new when it gets old. We will edit your report and get it back to you for approval if there are any significant changes or additions.
- Thank you! Without contributions like yours, *IEEE Security & Privacy* could not continue.