Reflections

Non-Asimovian Robots

Robert Silverberg

The assignment is to compare and contrast:

A) Isaac Asimov's Three Laws of Robotics, which he devised in December 1940. Actually, a generalized version of them was suggested to him by the great editor John W. Campbell, Jr., but it was Asimov who formulated the laws in explicit terms for a short story called "Runaround," published in 1942, one of a long series of stories and novels in which he would explore the social and technical consequences of the development of robots with quasi-human intelligence—

"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 the 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."

B) This passage from Philip K. Dick's novella, "Second Variety," first published in the May 1953 issue of *Space Science Fiction*—a story in which, after a nuclear war between the United States and the Soviet Union, the American survivors, based on the Moon, are attacking the Russian victors by means of deadly little remote-controlled robots called "claws" designed to infiltrate the Soviet bases and kill any troops they encounter—

"Across the ground something small and metallic came, flashing in the dull sunlight of mid-day. A metal sphere. It raced up the hill after the Russians, its treads flying. It was small, one of the baby ones. Its claws were out, two razor projections spinning in a blur of white steel. The Russian heard it. He turned instantly, firing. The sphere dissolved into particles. But already a second had emerged and was following the first. The Russian fired again.

"A third sphere leaped up the Russian's legs, clicking and whirring. It jumped to the shoulder. The spinning blades disappeared into the Russian's throat. . . .

"The claws weren't like other weapons. They were *alive,* from any practical standpoint, whether the Governments wanted to admit it or not. They were not machines. They were living things, spinning, creeping, snaking themselves up suddenly from the gray ash and darting toward the man, climbing up him, rushing for his throat. And that was what they had been designed to do. Their job."

The Dick story goes on to reveal that these lethal robots have established underground factories where they are now building later-model robots outwardly indistinguishable from human beings; and, as the story ends, we are told that "they were already beginning to design weapons to use against each other."

Isaac Asimov and Philip K. Dick were two of the most profound thinkers science fiction ever had. Yet their concepts of what a world equipped with intelligent robots would be like were wholly antithetical. Asimov, the humanist, grew up in the 1930s reading stories in which robots were diabolical menaces, essentially plantation hands who ultimately turned against their human masters—how many such stories were called "Revolt of the Robots" or something similar?—and, bored with such melodramatic fare, yearned "to write a robot story about a robot that was wisely used, that was not dangerous, and that did the job it was supposed to do." He wrote some stories in which the robots fulfilled those criteria; and by the time he had written the

fourth of them, John Campbell had helped him work out the eventually famous Three Laws of Robotics that would govern the world of the long series of robot stories Asimov was to write in the next four decades.

The kind of story Asimov wrote was intellectually based, avoiding violence and emotional outbursts, but he was a shrewd storyteller who knew that fiction needs conflict of some sort to support the plot. So the Asimov robot stories revolved around problems posed by the Three Laws, either when a robot found itself caught in a situation that created a logical contradiction between two of the laws, or when some unscrupulous manufacturer tried to find a loophole that would circumvent them altogether. The solution to the problem, when it came, was invariably an ingenious response to the challenge to the Laws. But Phil Dick, whose technophobic vision of the world was altogether darker than Asimov's technophilic one, chose entirely to ignore the Laws in his own numerous robot stories. He knew the Asimov stories, of course; everybody who read or wrote science fiction back then knew them. Dick, though, doubted that the coming age of robotics was likely to allow the niceties of moral imperatives to obscure the military or commercial value of artificial intelligence, and so he gave us the sinister war-machines of "Second Variety," the murderous synthetic entities of *Do Androids Dream of Electric Sheep?,* and many another tale of robots that are not just dangerous but downright terrifying.

And now we are moving swiftly into the robotic era that the twentieth-century SF writers foresaw: but is it the robots of Asimov or those of Dick that are coming among us?

Most of them, of course, are benign servants. www.Amazon.com has several thousand robots already at work in its warehouses, stubby little orange machines that carry shelves full of merchandise to human employees who select items from them and put them aboard conveyor belts that carry them along to workers who box and label them for shipment. (So far, the pickers and boxers are human, but I wouldn't be surprised to hear that Amazon has plans for changing that in the not too distant future.) In hospitals today, robot arms are used in surgical procedures too delicate and difficult for mere mortal doctors to perform. A hotel in Cupertino, California, the heart of Silicon Valley, uses a robot to deliver items requested by guests, moving unerringly through the hallways and calling the guests on their room phones to notify them of their arrival when the robot reaches the proper door. And, also in the San Francisco Bay Area, www.Google.com is currently testing driverless automobiles that it hopes one day will carry relaxed and cheerful passengers back and forth on our busy freeways while the riders sit back and amuse themselves with video games or even books. (As a resident of the area, I am not looking forward to this.) And in a million other ways, here in the second decade of the twenty-first century, machines are doing work that once was done by human beings, while hardly ever running amok.

But the drone that hovers over Afghanistan or Yemen and sends a missile into some terrorist group's headquarters is a kind of robot too, and although these are currently controlled by human operators thousands of miles away, it is not hard to foretell a time when a drone will be programmed to seek out some particular malefactor and use its own judgment about whether it has found the right target. Such an autonomous killer drone is not far in the future, and that will be a Dickian future, for such robots are as non-Asimovian as a robot can be. What if the target turns out to be hiding in a school or a hospital, though? Will the self-guided drone release its missile even so? Ronald Arkin, a roboticist at Georgia Tech, has been asked by the Pentagon to develop an "ethical adapter" that would permit a military robot to evaluate the destruction likely to result from the use of its weapons, so that when a built-in guilt level is reached, the robot will withhold its fire. This is not exactly Asimov's

First Law, but it does, at least, add a degree of moral compunction to the robot's array of capabilities.

The transition of robot technology from the pages of science fiction magazines to the real world brings real-world complexities that would have required Isaac to examine his Three Laws in new ways. Consider those driverless cars that Google hopes to set loose among us. A Google car finds itself on a collision course with one of the old-fashioned kind. It can avoid the collision and save all the passengers by swerving; but then it will hit a pedestrian. What does the First Law say about that? Suppose it has to choose between swerving into the car on its right that has four passengers, or the one on its left that has only one? If a driverless car is unoccupied at the moment, driving along the side of a steep cliff, and a head-on collision with an oncoming human-occupied car is imminent, it can save the situation by letting itself go over the edge, using the Third Law as its justification. But what happens at a four-way intersection governed by stop signs if the other three drivers are human, operating by sometimes irrational human decision-making techniques, and the fourth is a Three Laws robot? The driverless car must take the first and third Asimov Laws into account in deciding when to move forward, while the three human drivers are guided by thinking of a very different sort, centering on the desire to get through that intersection ahead of the other guys.

The movie industry, of course, is inherently non-Asimovian, and science fiction films have long been filled with dangerous, malevolent, and downright malicious robots. Asimov's good friend Arthur C. Clarke gave us *2001*'s HAL-9000, who responds to a threat by a spaceship's human crew by trying to kill all of them, and very nearly succeeds. The Terminator movies show the self-aware Skynet network planning to destroy humanity. A cowboy robot in *Westworld* malfunctions and starts aiming its six-shooter at its human companions. The list goes on and on.

Even Isaac Asimov, ironically, fell victim to Hollywood's bad-robot syndrome, in a movie called *I, Robot* that was more or less based on his own first collection of robot stories. Fortunately for Isaac, the film, quite a terrible one, was made after his death; it certainly would have sent his blood pressure soaring if he had lived to see it. The Three Laws are in it, yes, but a new robot model, the NS–5, has figured out how to countermand them. Not only are they capable of attacking human beings, but the aggressive NS–5s have set about dismantling the harmless older models and building more of their own kind, leading up to a frenzied and explosive robot takeover of Chicago—complete with all the wild Revolt of the Robots melodrama that had led Isaac Asimov to postulate the Three Laws in the first place in reaction to the anti-scientific attitudes that such melodramatics inspire. Ultimately the supremacy of the Three Laws is restored, but not before a lot of sound and fury has been unleashed. Poor Isaac!

And here we are, moving on into a glorious new century where robots will soon surround us on every side. Will they be Asimovian robots or Dickian robots? I know what I think: that a sincere effort will be made to build something akin to the Three Laws into any mechanism that has the capacity to endanger human beings, and that—sooner or later—the demands of commerce or the inexorable needs of the military will find a way of overriding those laws. Welcome to your grand and perilous robotic future, folks.