My Robots

Isaac Asimov

I wrote my first robot story, “Robbie,” in May of 1939, when I was only nineteen years old.

What made it different from robot stories that had been written earlier was that I was determined not to make my robots symbols. They were not to be symbols of humanity’s overweening arrogance. They were not to be examples of human ambitions trespassing on the domain of the Almighty. They were not to be a new Tower of Babel requiring punishment.

Nor were the robots to be symbols of minority groups. They were not to be pathetic creatures that were unfairly persecuted so that I could make Aesopic statements about Jews, Blacks or any other mistreated members of society. Naturally, I was bitterly opposed to such mistreatment and I made that plain in numerous stories and essays—but not in my robot stories.

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.

Since I began writing my robot stories in 1939, I did not mention computerization in their connection. The electronic computer had not yet been invented and I did not foresee it. I did foresee, however, that the brain had to be electronic in some fashion. However, “electronic” didn’t seem futuristic enough. The positron—a subatomic particle exactly like the electron but of opposite electric charge—had been discovered only four years before I wrote my first robot story. It sounded very science fictional indeed, so I gave my robots “positronic brains” and imagined their thoughts to consist of flashing streams of positrons, coming into existence, then going out of existence almost immediately. These stories that I wrote were therefore called “the positronic robot series,” but there was no greater significance than what I have just described to the use of positrons rather than electrons.

At first, I did not bother actually systematizing, or putting into words, just what the safeguards were that I imagined to be built into my robots. From the very start, though, since I wasn’t going to have it possible for a robot to kill its creator, I had to stress that robots could not harm human beings; that this was an ingrained part of the makeup of their positronic brains.

Thus, in the very first printed version of “Robbie,” I had a character refer to a robot as follows: “He just can’t help being faithful and loving and kind. He’s a machine, made so.”

After writing “Robbie,” which John Campbell, of Astounding Science Fiction, rejected, I went on to other robot stories which Campbell accepted. On December 23, 1940, I came to him with an idea for a mind-reading robot (which later became “Liar!”) and John was dissatisfied with my explanations of why the robot behaved as it did. He wanted the safeguard specified precisely so that we could understand the robot. Together, then, we worked out what came to be known as the “Three Laws of Robotics. “ The concept was mine, for it was obtained out of the stories I had already written, but the actual wording (if I remember correctly) was beaten out then and there by the two of us.

The Three Laws were logical and made sense. To begin with, there was the question of safety, which had been foremost in my mind when I began to write stories about my robots. What’s more I was aware of the fact that even without actively attempting to do harm, one could quietly, by doing nothing, allow harm to come. What was in my mind was Arthur Hugh Clough’s cynical “The Latest Decalog,” in which the Ten Commandments ate rewritten in deeply satirical Machiavellian fashion. The one item most frequently quoted is: “Thou shalt not kill, but needst not strive / Officiously to keep alive.”

For that reason I insisted that the First Law (safety) had to be in two parts and it came out this way:

1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.

Having got that out of the way, we had to pass on to the second law (service). Naturally, in giving the robot the built-in necessity to follow orders, you couldn’t forfeit the overall concern of safety. The Second Law had to read as follows, then:

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

And finally, we had to have a third law (prudence). A robot was bound to be an expensive machine and it must not needlessly be damaged or destroyed. Naturally, this must not be used as a way of compromising either safety or service. The Third Law, therefore, had to read as follows:

3. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Laws.

Of course, these laws are expressed in words, which is an imperfection. In the positronic brain, they are competing positronic potentials that are best expressed in terms of advanced mathematics (which is well beyond my ken, I assure you). However, even so, there are clear ambiguities. What constitutes “harm” to a human being? Must a robot obey orders given it by a child, by a madman, by a malevolent human being? Must a robot give up its own expensive and useful existence to prevent a trivial harm to an unimportant human being? What is trivial and what is unimportant?

These ambiguities are not shortcomings as far as a writer is concerned. If the Three Laws were perfect and unambiguous there would be no room for stories. It is in the nooks and crannies of the ambiguities that all one’s plots can lodge, and which provide a foundation, if you’ll excuse the pun, for Robot City.

I did not specifically state the Three Laws in words in “Liar!” which appeared in the May 1941 Astounding. I did do so, however, in my next robot story, “Runaround,” which appeared in the March 1942 Astounding. In that issue on line seven of page one hundred, I have a character say, “Now, look, let’s start with the three fundamental Rules of Robotics,” and I then quote them. That incidentally, as far as I or anyone else has been able to tell, represents the first appearance in print of the word “robotics”—which, apparently, I invented.

Since then, I have never had occasion, over a period of over forty years during which I wrote many stories and novels dealing with robots, to be forced to modify the Three Laws. However, as time passed, and as my robots advanced in complexity and versatility, I did feel that they would have to reach for something still higher. Thus, in Robots and Empire, a novel published by Doubleday in 1985, I talked about the possibility that a sufficiently advanced robot might feel it necessary to consider the prevention of harm to humanity generally as taking precedence over the prevention of harm to an individual. This I called the “Zeroth Law of Robotics,” but I’m still working on that.

My invention of the Three Laws of Robotics is probably my most important contribution to science fiction. They are widely quoted outside the field, and no history of robotics could possibly be complete without mention of the Three Laws. In 1985, John Wiley and Sons published a huge tome, Handbook of Industrial Robotics, edited by Shimon Y. Nof, and, at the editor’s request, I wrote an introduction concerning the Three Laws.

Now it is understood that science fiction writers generally have created a pool of ideas that form a common stock into which all writers can dip. For that reason, I have never objected to other writers who have used robots that obey the Three Laws. I have, rather, been flattered and, honestly, modem science fictional robots can scarcely appear without those Laws.

However, I have firmly resisted the actual quotation of the Three Laws by any other writer. Take the Laws for granted, is my attitude in this matter, but don’t recite them. The concepts are everyone’s but the words are mine.