The Laws of Humanics

Isaac Asimov

My first three robot novels were, essentially, murder mysteries, with Elijah Baley as the detective. Of these first three, the second novel, The Naked Sun, was a locked-room mystery, in the sense that the murdered person was found with no weapon on the site and yet no weapon could have been removed either.

I managed to produce a satisfactory solution but I did not do that sort of thing again.

The fourth robot novel, Robots and Empire, was not primarily a murder mystery. Elijah Baley had died a natural death at a good, old age, the book veered toward the Foundation universe so that it was clear that both my notable series, the Robot series and the Foundation series, were going to be fused into a broader whole. (No, I didn’t do this for some arbitrary reason. The necessities arising out of writing sequels in the 1980s to tales originally written in the 19408 and 1950s forced my hand.)

In Robots and Empire, my robot character, Giskard, of whom I was very fond, began to concern himself with “the Laws of Humanics,” which, I indicated, might eventually serve as the basis for the science of psychohistory, which plays such a large role in the Foundation series.

Strictly speaking, the Laws of Humanics should be a description, in concise form, of how human beings actually behave. No such description exists, of course. Even psychologists, who study the matter scientifically (at least, I hope they do) cannot present any “laws” but can only make lengthy and diffuse descriptions of what people seem to do. And none of them are prescriptive. When a psychologist says that people respond in this way to a stimulus of that sort, he merely means that some do at some times. Others may do it at other times, or may not do it at all.

If we have to wait for actual laws prescribing human behavior in order to establish psychohistory (and surely we must) then I suppose we will have to wait a long time.

Well, then, what are we going to do about the Laws of Humanics? I suppose what we can do is to start in a very small way, and then later slowly build it up, if we can.

Thus, in Robots and Empire, it is a robot, Giskard, who raises the question of the Laws of Humanics. Being a robot, he must view everything from the standpoint of the Three Laws of Robotics—these robotic laws being truly prescriptive, since robots are forced to obey them and cannot disobey them.

The Three Laws of Robotics are:

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.

Well, then, it seems to me that a robot could not help but think that human beings ought to behave in such a way as to make it easier for robots to obey those laws.

In fact, it seems to me that ethical human beings should be as anxious to make life easier for robots as the robots themselves would. I took up this matter in my story “The Bicentennial Man,” which was published in 1976. In it, I had a human character say in part:

“If a man has the right to give a robot any order that does not involve harm to a human being, he should have the decency never to give a robot any order that involves harm to a robot, unless human safety absolutely requires it. With great power goes great responsibility, and if the robots have Three Laws to protect men, is it too much to ask that men have a law or two to protect robots?”

For instance, the First Law is in two parts. The first part, “A robot may not injure a human being,” is absolute and nothing need be done about that. The second part, “or, through inaction, allow a human being to come to harm,” leaves things open a bit. A human being might be about to come to harm because of some event involving an inanimate object. A heavy weight might be likely to fall upon him, or he may slip and be about to fall into a lake, or anyone of uncountable other misadventures of the sort may be involved. Here the robot simply must try to rescue the human being; pull him from under, steady him on his feet and so on. Or a human being might be threatened by some form of life other than human—a lion, for instance—and the robot must come to his defense.

But what if harm to a human being is threatened by the action of another human being? There a robot must decide what to do. Can he save one human being without harming the other? Or if there must be harm, what course of action must he pursue to make it minimal?

It would be a lot easier for the robot, if human beings were as concerned about the welfare of human beings, as robots are expected to be. And, indeed, any reasonable human code of ethics would instruct human beings to care for each other and to do no harm to each other. Which is, after all, the mandate that humans gave robots. Therefore the First Law of Humanics from the robots’ standpoint is:

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

If this law is carried through, the robot will be left guarding the human being from misadventures with inanimate objects and with non-human life, something which poses no ethical dilemmas for it. Of course, the robot must still guard against harm done a human being unwittingly by another human being. It must also stand ready to come to the aid of a threatened human being, if another human being on the scene simply cannot get to the scene of action quickly enough. But then, even a robot may unwittingly harm a human being, and even a robot may not be fast enough to get to the scene of action in time or skilled enough to take the necessary action. Nothing is perfect.

That brings us to the Second Law of Robotics, which compels a robot to obey all orders given it by human beings except where such orders would conflict with the First Law. This means that human beings can give robots any order without limitation as long as it does not involve harm to a human being.

But then a human being might order a robot to do something impossible, or give it an order that might involve a robot in a dilemma that would do damage to its brain. Thus, in my short story “Liar!,” published in 1940, I had a human being deliberately put a robot into a dilemma where its brain burnt out and ceased to function.

We might even imagine that as a robot becomes more intelligent and self-aware, its brain might become sensitive enough to undergo harm if it were forced to do something needlessly embarrassing or undignified. Consequently, the Second Law of Humanics would be:

2—A human being must give orders to a robot that preserve robotic existence, unless such orders cause harm or discomfort to human beings.

The Third Law of Robotics is designed to protect the robot, but from the robotic view it can be seen that it does not go far enough. The robot must sacrifice its existence if the First or Second Law makes that necessary. Where the First Law is concerned, there can be no argument. A robot must give up its existence if that is the only way it can avoid doing harm to a human being or can prevent harm from coming to a human being. If we admit the innate superiority of any human being to any robot (which is something I am a little reluctant to admit, actually), then this is inevitable.

On the other hand, must a robot give up its existence merely in obedience to an order that might be trivial, or even malicious? In “The Bicentennial Man,” I have some hoodlums deliberately order a robot to take itself apart for the fun of watching that happen. The Third Law of Humanics must therefore be:

3—A human being must not harm a robot, or, through inaction, allow a robot to come to harm, unless such harm is needed to keep a human being from harm or to allow a vital order to be carried out.

Of course, we cannot enforce these laws as we can the Robotic Laws. We cannot design human brains as we design robot brains. It is, however, a beginning, and I honestly think that if we are to have power over intelligent robots, we must feel a corresponding responsibility for them, as the human character in my story “The Bicentennial Man” said.