

Religion Today  
April 15, 2015  
**“Robot Morality”**  
Paul V.M. Flesher

In research worthy of science fiction writer Isaac Asimov's "I, Robot," Bertram Malle is working to design a moral robot. Malle is the co-director of Brown University's Humanity-Centered Robotics Initiative, and his approach is to create a robot that can learn moral behavior from the people around it. Ideally, you would surround the robot with morally good people, and the robot would learn ethical beliefs and behavior from them.

Like a child and its parents, the robot then would be taught morality and behavior by the people looking after it. Of course, there would be no need to limit the teachers to just two people. Once beyond the basics, robots could even crowd-source their ethical education. When two principles it learns come into conflict, the robot could seek guidance and feedback from those it knows.

But what happens if the robot falls in with the wrong crowd? Perhaps the robot gets stolen by a criminal gang that teaches it how to be a thief or a murderer.

To avoid such a scenario, the robot should be equipped with a set of core rules that would guide its learning. Like Asimov's "Three Laws of Robotics," the guidelines would direct the robot away from doing harm and evil and toward doing good. The key question, then, is what are those rules?

Malle indicates these rules would need to include the prevention of harm to humans, like Asimov's Law 1 "A robot may not injure a human being or, through inaction, allow a human being to come to harm," as well as guidance concerning the politeness and respect required for smooth human interactions.

Another rule that would be needed is to treat all people the same, that is, according to the same ethical principles and behavior. As Malle puts it, "we can equip robots with an unwavering prosocial orientation. As a result, they will follow moral norms more consistently than humans do, because they don't see them in conflict, like humans do, with their own selfish needs."

The problem with human morality Malle identifies here is the selfishness of each individual. Selfishness often prevents humans from doing what they consider to be the morally correct act. Robots would not be diverted from moral behavior by selfishness because they lack a self. They have as much self-awareness as a TV or a refrigerator. They would be a moral machine, always behaving ethically, without any personal needs or desires to sidetrack them.

But there is a second problem with human morality, namely, to whom should ethical behavior apply? Humans are always joining with other people in groups, and people often treat members of these groups differently from those who do not belong.

Family members treat each other differently from the way they treat non-family members. Friends behave differently toward each other than toward mere acquaintances.

We conduct our relations with members of our religious organization differently from those who do not belong, or more importantly, from those who disagree with our religion. The fracas in Indiana about religious freedom and discrimination against gays is a case in point.

Other groups affect our behavior toward others. During an election, we behave differently toward members of different political parties.

Some people treat members of certain racial groups or ethnic groups different from those of our own. Just think about our current national argument over white police shooting black citizens, or the problems surrounding Hispanic immigration.

Once robots are programmed with the rule to treat all people the same, without regard to their group membership, these problems would be avoided. Since robots have no more self than a pickup truck, the human tendency to identify their self with groups would not take place. Robots would have no reason to treat Hispanics or Asians differently from whites. They would not behave toward evangelical Christians with one set of moral standards, toward Catholics with another and toward Muslims with a third.

In other words, robots would be more moral than human beings. Their ability to perform in a morally consistent manner toward everyone they meet would be superior to our own.

Of course, research into robotics has not yet reached the ability to program robots in this way, but scientists like Malle are working toward that goal. It is sobering to think, however, that robots could outperform humans not only in raw calculating and thinking power, but also in terms of ethical behavior.

Note: This essay draws from "How to Raise a Moral Robot," by Bertram Malle, *livescience*, April 2, 2015 (<http://www.livescience.com/50349-how-to-raise-a-moral-robot.html>).