Teaching Robots Right from Wrong

You're rushing across the school parking area to get to your first class on time when you notice a friend is in trouble. She's texting and listening to music on her earphones. Unaware that she's also heading straight for a gaping hole in the pavement. What do you do?
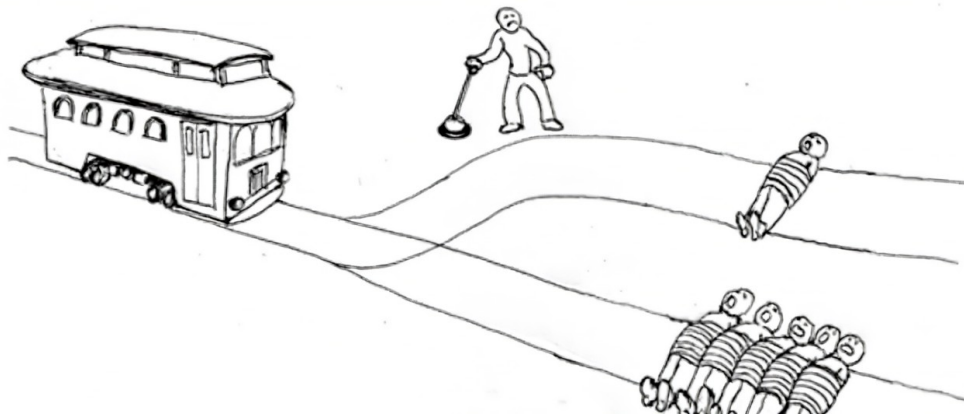
To figure out the best solution, such a decision balances the effects of your choice. It's an easy decision. You don't even have to think hard about it. You make such choices all the time. But what about robots? Can they make such choices? Should a robot stop your friend from falling into the hole? Could it?

Not today's robots. They simply aren't smart enough to even realise when someone is in danger. Soon, they might be. Yet without some rules to follow, a robot wouldn't know the best choice to make. So robot developers are turning to philosophy. Called ethics, it's a field in which people study differences between right and wrong. And with it, they are starting to develop robots that can make basic ethical decisions.

Such research should help robots of the future figure out the best action to take when there are competing choices. This ethical behavior may just become part of their programming. That will allow them to interact

with people in safe, predictable ways. In time, robots may actually begin to understand the difference between right and wrong. The most famous set of rules for robots comes not from research but from a science fiction story by Isaac Asimov. "Runaround," published in 1942, features two men and Robot SPD-13, nicknamed "Speedy." They're sent to the planet Mercury in the year 2015. Speedy is programmed with three basic rules: 1) A robot can't hurt a person or, through inaction, allow a person to get hurt. 2) A robot must obey people, as long as this doesn't break the first law. 3) A robot must protect itself, as long as this doesn't break the first two laws. In later robot stories, Asimov added a "zeroth" law: A robot can't harm humanity or, through inaction, allow harm to humanity. Asimov's rules sound good. But the story shows that such simple rules may not be enough. These rules would certainly compel a robot to rescue your friend. But they wouldn't help a robot decide what to do if two people were about to fall and it could only save one. The robot also wouldn't try to rescue a kitten. It's very difficult to write a set of rules that will apply in all possible situations. For this reason, some scientists instead build robots with the ability to learn ethical behavior.

Kathryn Hulick, Science News For Students. April 20th, 2017

The trolley problem.