I, Car: The Four Laws of Robotic Cars

> *"let's start with the three fundamental Rules of Robotics — the three rules that are built most deeply into a robot's positronic brain."*
>
> Isaac Asimov

The science fiction writer Isaac Asimov made reference to the "Three laws of Robotics" in several of his stories, which are a kind of safeguard for humans in the future, when robots could become intelligent and self-aware. In the compilation "I, Robot" Asimov provides interesting examples of the troubles for humans and robots when intelligent behavior leads to conflicting scenarios. The three Laws 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 to 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 Laws.

General purpose robots are not still intelligent, nor will they be in the next decades. However, autonomous cars are already driving the streets in several cities, so that the issue of legal liability is starting to come up. Typically, autonomous cars have some sort of insurance and they are not really allowed to go driverless in the cities. They still need a human as supervisor, who must be ready to take control in case of an emergency. However, maybe it is time to formulate some rules that robotic cars must follow when they become really driverless. In the spirit of Asimov these rules would be:

1. A car may not injure a human being or, through inaction, allow a human being to come to harm.
2. A car must obey the traffic rules, except when they would conflict with the First Law.
3. A car must obey the orders given to it by human beings, except where such orders would conflict with the First or Second Laws.
4. A car must protect its own existence as long as such protection does not conflict with the First, Second or Third Laws.

Notice that the main problem with Asimov's Three Laws is that they do no deal with the environment. Typically, there are rules and restrictions which must be respected, especially in traffic. Robots do not act in a regulatory vacuum. Therefore, we need an additional Law, conveniently placed after the original First Law. This new Second Law constraints all car activity to fall within the bounds of existing traffic laws. For example, a human cannot order a car to go in the wrong traffic direction, even when the street is empty, nor can make it disregard a red light when nobody is looking. If property damage is not allowed in the traffic rules, a human cannot order a car to self-destroy.

With these robotic laws an autonomous car would be pro-active: it would change the parking space in case of high-water, or it would even try to save a human by blocking a

falling boulder, for example. Of course, Asimov could easily come out with a few examples of everything that can go wrong even with these four rules.

Raul Rojas
Freie Universität Berlin