

Should Self-Driving Cars be bound by Isaac Asimov's Three Laws of Robotics?

Pradhyumnaa Ganapathi Subramanian

July 2019

1 Introduction

Autonomous cars are on the rise and since they are considered intelligent systems, an interesting question arises. Should an autonomous car be bound by the Three Laws of Robotics?

2 Connection between the definition of a robot and an autonomous car

According to The Institute of Electrical and Electronics Engineers, the definition of a Robot is an autonomous machine capable of sensing its environment, carrying out computations to make decisions, and performing actions in the real world.

An "Autonomous" car could be categorized as an intelligent control system that combines the functions of variety of sensors to deduce information about the environment. The autonomous car also makes computations to make decisions on queries such as "What is the shortest path to the destination?" and also drives to reach the specified destination. Since the actions of an autonomous car coincide with the definition of a robot, an autonomous car could be considered a robot.

3 Isaac Asimov's Three Laws of Robotics

The Three Laws of Robotics are a set of rules which all robots are expected to follow.

1. **The First Law:** A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. **The Second Law:** A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. **The Third Law:** A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Since an autonomous car could be considered a robot, an interesting question arises. Should an Autonomous Car be expected to obey the 3 Laws of Robotics and more importantly, can an Autonomous Car be made to follow the 3 Laws of Robotics?

4 Cosmetic Alterations to the Laws

Slight modifications to the 'law' such as making the machine to consider other factors beyond those of which fall under the 3 Laws of Robotics (For Example, considering the societal or economic status of the human being involved) will be reviewed.

A fitting example of this would be the MoralMachine study conducted by Massachusetts Institute of Technology. In this experiment, a quiz was devised with questions that would make the quiz-taker select in which scenarios the passengers of the car should be saved and in which the pedestrians should be saved.

Actions	Probability
Preferring Inaction	0.47
Sparing Pedestrians	0.57
Sparing Females	0.46
Sparing the Fit	0.57
Sparing the Lawful	0.44
Sparing the Higher Status	0.49
Sparing the Younger	0.56
Sparing More	0.51
Sparing Humans	0.66

Table 1: Results of the MoralMachine Test (World Average)

As stated in the table, 47 percent of the people picked to not act in the given scenario. This would mean that the robot, the Autonomous Car then would have to injure a human (either the passenger or the pedestrian) being which violates the First Law of Robotics.

One interesting thing to note is according to the results, people only chose to save the lawful 44 percent of the time. From an ethical and lawful point of view, this is absurd. The pedestrian in this case, was a law-abiding citizen who was just at the wrong place at the wrong time.

This could create a "deadlock" as the Autonomous Car could be considered a liability from the passengers' point of view as there is a probability the car chooses to crash against a barrier, possibly killing the passengers and there is a probability that the car choosing to crash into a law abiding pedestrian.

In 2018, an Uber Self Driving car killed a pedestrian and made the news. It was stated that around 94 percent of the accidents were the driver's error but

then who is to be held accountable for around 3 percent of the accidents? In that situation, who is to be blamed? The car, the manufacturer who designed the system's artificial intelligence or the passenger? (Assuming the remaining 3 percent of the accidents were caused by the system's inability to make a decision during the situation.)

No autonomous vehicle is truly autonomous since it is impossible to expose the system to all possible scenarios which would be used to train and/or test the system as this process would be very expensive for practical implementation.

5 Conclusion

The first law would not be satisfied in any situation where the car would have to make a decision on whether to crash onto a barricade, possibly killing passenger or crash into the pedestrian, saving the passenger but possibly killing the pedestrian.

Even when one of the laws are not being followed, it is not possible to make an autonomous vehicle obey to the 3 laws of robotics. Until and unless some alterations are made to the laws to fit the purpose of an autonomous vehicle, the laws can not be obeyed by an autonomous car.

6 References

1.What is a Robot? <https://robots.ieee.org/learn/> 2.Moral Machine Graph. <http://moralmachineresults.scalablecoop.org/> 3.Uber Self Driving Car Incident <https://www.economist.com/the-economist-explains/2018/05/29/why-ubers-self-driving-car-killed-a-pedestrian>