

# The Ethical Considerations Of Self-driving Cars

Maryo Botros

mbotros@oxy.edu

Occidental College

## Abstract

This is the abstract

## 1 Introduction

Although they have not been adopted on a large scale, self-driving cars already exist and as with any new technology, they pose some ethical dilemmas.

## 2 The Trolley Problem

The main ethical conflict with autonomous vehicles involves a decision between prioritizing the interests of passengers (arriving quickly and efficiently) or the interests of the community as a whole (making sure roads are safe for everyone using them) [1]. Utilitarianism is the ethical approach that sides with the community interest and it is “the greatest amount of good for the greatest number of people” [1]. A utilitarian approach to crashes would mean minimizing overall casualties in any manner necessary, including sacrificing the passenger of the autonomous vehicle to save the lives of more pedestrians. This invokes the famous trolley problem, a thought experiment in ethics about a fictional scenario in which an onlooker has the choice to save 5 people in danger of being hit by a trolley, by diverting the trolley to kill just 1 person (Merriam-Webster). Studies have shown that an overwhelming number of people find this approach to be the most moral; one study in particular concluded that 76% of people thought it more morally good to sacrifice one passenger to save ten pedestrians than vice versa [1]. On the other hand, the studies also found that participants were more likely to buy an autonomous car programmed to save its own passengers over any number of pedestrians, when offered the hypothetical choice between that and a utilitarian car [1].

These results reveal that this ethical dilemma is also a social one: everyone wants others to behave in a way that leads to the best global outcome, but they are also hypocritically not motivated to practice that behavior themselves. Additionally, the utilitarian approach to accidents would theoretically save the maximum number of people, but consumers don’t seem to want to buy these vehicles and under-

standably so because they do not guarantee the safety of the passengers and do not necessarily prioritize the passenger’s safety over the safety of others. “Counterintuitively, this may mean that non-utilitarian AVs would be ethically better, since people would be more likely to buy them, resulting in a greater number of lives saved overall simply by having more autonomous vehicles on the road” [1]. Furthermore, a utilitarian approach to crashes does not take into account more complex factors, including the social value of victims, such as their age or ethnicity, and whether it is morally worse to actively kill rather than kill through inaction [1]. Any algorithm that makes decisions based on social value would need to be supported by a scale of social value, and obviously, not everyone has the same idea about the relative value of different lives [1]. Would a pregnant woman have more social value than a child, for example? Not only is this a difficult dilemma to solve, but this would also mean that the vehicle would need to implement this sort of decision-making. Autonomous vehicles would have to be able to identify and categorize the people involved in each crash based on social factors that may well not be visible on their person, which is beyond the realm of current sensor technology. Sensors would not be able to detect if a person is a murderer versus a school teacher who has never committed a crime. The question of killing by inaction is slightly more straightforward and can be addressed with the Principle of Double-Effect, developed by Christian philosophers. This principle would advise never to swerve an autonomous vehicle regardless of the number of pedestrians that would be killed, as doing so would result in intentional harm to others, while keeping the Autonomous vehicle on its original path would be considered “merely foreseen harm” (more ethically permissible than the former) [1]. However, realistically, it would likely be difficult for autonomous vehicle companies to justify their cars killing thirty pedestrians to avoid swerving and hitting one.

Another ethical framework, called the Rights Approach, could be considered the opposite of the Utilitarian Approach. It skews more towards the interests of the individual, which could be the passengers of the autonomous car or each pedestrian on the road. One interpretation of this approach indicates that since the main goal of an autonomous vehicle is to move passengers around safely, the passenger has a right to life [1]. But the lives of everyone should be

considered, since the autonomous vehicle manufacturer has decided to make a “dangerous machine” and the passengers have decided to ride in the dangerous machine, they both owe pedestrians a duty of care to safety [1]. Since the passengers already understand the risks associated with riding in an autonomous vehicle, it would be more morally right to prioritize the lives of others on the road since they are not taking that same risk. However, if autonomous vehicle manufacturers programmed their cars this way, consumer interest would likely drop immensely, and since that means fewer autonomous vehicles on the road, this might result in a net loss of life due to more casualties from manned vehicles (same as the drawback to the utilitarian approach) [1]. So, both the Utilitarian Approach and the Rights approach seem to result in the catch-22 that programming autonomous vehicles with the more morally correct approach will end in more loss of life overall due to consumers’ selfish, yet understandable desire to protect themselves above others [1]. There does not seem to be a better solution to this problem other than creating a car that would be perfect and would not have to decide between saving the lives of passengers or others on the road, but this would simply not be possible as encountering uncertainty on the road is inevitable. Another solution would be to have a passenger monitoring the car as it autonomously navigates its environment. Again, however, this would defeat the purpose of an autonomous car as now it would require human interaction.

### 3 Allocating Responsibility

One of the major ethical dilemmas posed by self-driving cars is the allocation of responsibility. Ideally, self-driving cars should never be involved in any accidents, as that is one of the things they are meant to prevent. However, accidents have occurred because of self-driving car technology and sensors can sometimes be faulty when perceiving the environment, so an accident is always a possibility, and understanding who is responsible for the accident can be an issue. Just about all self-driving cars on the road today require that a driver be in the driver’s seat and this person has just as much responsibility as any other driver on the road. However, this is only a temporary solution as self-driving cars are meant to be completely autonomous and all people in the car will be considered proper passengers. When we reach the point where cars are completely autonomous and the computer processing artificial intelligence is the entity that is “driving,” who is responsible for the safety of the passengers in the car as well as the safety of others traveling and walking on the same roads? Would it be the vehicle owner, vehicle manufacturer, passengers, or someone else?

Two major forms of responsibility can be used to analyze the issues with allocating responsibility when it comes to self-driving cars: task responsibility, commonly referred

to as “forwards-looking responsibility,” and blame responsibility, often called “backwards-looking” responsibility. Having a task responsibility means being obliged to do something. Having a blame responsibility means that one is to be blamed if something goes wrong [2]. Blame responsibility is often associated with punishments or with duties to compensate. Our responsibility ascriptions will certainly change when driverless cars are introduced. Users of fully automated vehicles have no control over the vehicle, other than their choice of a destination, so it would be difficult to hold them responsible either for safety (task responsibility) or for accidents (blame responsibility) because we cannot hold people responsible for something they have no control over [2].

There are three main alternatives for what can be done instead. First, we can hold other people responsible instead, such as the vehicle manufacturers and the people responsible for the road system (including the communication and coordination systems used to guide the vehicles). The second option is to hold the artificial intelligence built into the vehicles responsible. The third is to treat traffic accidents in the same way as natural accidents such as tsunamis, strokes of lightning, and earthquakes, for which no one is held responsible [2]. Although the future is always difficult to predict, the first option is by far the most probable one. Previous experience shows that this is how responsibility is assigned when a human is replaced by an automatic system. For instance, if an aviation accident unfolds after the pilot turned on the autopilot, we do not blame the artificial intelligence that took over the flight, and neither do we treat the failure as a natural event. Instead, we assign blame to those who directed the construction, testing, installation, service, and updating of artificial intelligence. Such an approach is not unknown in road traffic. With the way semi-autonomous vehicles work now, the approaches to traffic safety have had some success in achieving an analogous transfer of responsibility to vehicle and road system providers, although human drivers are still in place and these vehicles are not autonomous [2].

### 4 Data

In order for self-driving cars to work effectively, a lot of data needs to be collected and stored such as data from the car’s sensors. This data helps the cars get smarter in a way; as they collect more data, the cars gain a better understanding of their environment and therefore better at navigating the environment. The more data that is collected and processed, the more effective the car will be autonomously driving passengers around. Storing this data can pose some security risks, however, as this data includes the location of the cars and therefore the passenger’s locations. Although this information is vital for the success of the autonomous

system, some may consider the collection of this sort of data to be unethical because it can potentially end up in the wrong hands and if it does, it could be detrimental to a person's identity, finances, and livelihood.

Something else to consider when it comes to the data that is collected by autonomous cars is that the information gathered by the vehicle itself would be complemented by well-developed communication systems, meaning vehicle-to-vehicle communication as well as vehicle-to-road management communications and this would put even more data at risk. Vehicle-to-vehicle communication would allow for information about obstacles ahead to be obtained before they are registered by the car's own sensors. Furthermore, sensor or sensor interpretation errors can be detected by comparison with information from other cars or from the roadside [2]. If vehicle-to-road-management systems are interconnected on a large scale, then they can also be used for optimizing the traffic flow because they can provide updated local information on traffic and accessibility.

Handling person-related information and collecting and processing traffic information can give rise to potential privacy intrusions. Currently, private cars do not necessarily leave any electronic traces, but this would change when the number of autonomous vehicles on the road increases as these vehicles will depend on ge positioning transponders and possibly on centralized systems that keep track of each vehicle's planned route and destination and this information will be linked to the owner and possibly even the passengers [2].

Ge positioning can be a particularly concerning problem as it is highly sensitive. It can reveal information about individuals such as where they live and work and even membership in a religious or political organization, or even information about private relationships [2]. This sort of information can be life-threatening in extreme situations. Currently, people's web browsing data in the virtual space of the internet is used to tailor a massive flow of advertisements for them. With ge positioning, data on people's movements in real space can be used in a similar fashion. This would mean that sellers and rental providers of vehicles will have economic incentives to include an advertisement function over which they retain control so that they can sell space on it [2]. For instance, if a car was parked outside of a shopping mall, the owner or renter of the car would receive commercial messages from different companies that operate at the mall and these messages could be conveyed through the speakers, screens in the car, or other media connected with the person who owns or rents the car.

When an economic incentive becomes part of the equation, this may create other problems, such as interference with navigation directions, which could come in the form of the system not providing the fastest route to a desired destination. For example, if a person makes a general search

for a "grocery store," the navigation system may provide a grocery store brand that the vehicle owner has gone to before or has searched for on the web instead of the one that is the closest. Similar to the way people receive ads from previous sites they visited, as they search the internet, they may receive tailored results for navigation to different destinations.

Even if the communication between autonomous vehicles becomes efficient, most accidents will likely happen because of cars driven by humans and unprotected travelers such as pedestrians and cyclists. One solution to this would be for these non-autonomous vehicles and pedestrians to carry a transponder that communicates with the autonomous vehicles in order to avoid accidents [2]. Demand for transponders may arise to make them mandatory for certain types of vehicles or pedestrians and these personal transponders would give rise to the same privacy issues as ge positioning on autonomous vehicles [2].

## 5 Conclusion

In conclusion, the introduction of autonomous vehicles on the road gives rise to a plethora of ethical problems. It will change the way in which responsibility descriptions concerning accidents and traffic safety and responsibility will fall more on the constructors and maintainers of the autonomous vehicles, roads, and their communication systems. Autonomous vehicles will require the collection of extensive sensitive information in order to optimize navigation and this information can be misused. Additionally, autonomous vehicles may work best if others sharing the road use transponders and the use of these transponders will give rise to the same issues concerning privacy as the transponders used in the autonomous vehicles. Because of all of these reasons, it seems to be quite clear that autonomous cars pose too many risks and dangers and the costs far outweigh the benefits. If autonomous vehicles were to be adopted, there would need to be a lot more research conducted in order to address and solve the ethical problems associated with them.

## References

- [1] Ampe, Teagan. *Autonomous Accidents: The Ethics of Self-Driving Car Crashes*. 2020. URL: <https://vce.usc.edu/volume-4-issue-2/autonomous-accidents-the-ethics-of-self-driving-car-crashes/>.
- [2] Hansson. *Self-Driving Vehicles—an Ethical Overview*. 2021. URL: <https://doi.org/10.1007/s13347-021-00464-5>.