Stephen Kemp

EE129A; Fall 2018

November 6, 2018

Ethical Discussion of Self-Driving Cars

With level-4[[1]](#footnote-0) autonomous vehicles being on the immediate horizon, there are many engineering problems being solved to create this technology. These problems can be framed in terms of an objective goal, and a process by which the engineer maximizes a function of desired parameters to find a sufficiently optimal solution. More difficult to address are the moral and ethical problems that arise from realistic scenarios involving autonomous vehicles, especially relating to vehicle-pedestrian accidents, or system failure where people may be injured or killed. In these cases, there are ethical decisions that the engineer needs to make months or years in advance of the incident. Additionally, ethical liability for incidents and failures needs to be discussed; Who is at fault if the car malfunctions, or makes a decision resulting in someone’s death?

Because of the complexity of these moral issues the question must be addressed; should self-driving cars even be deployed in the first place? For the purposes of this essay, I will work under the assumption that self-driving cars will present a significant reduction in loss of life over human drivers[[2]](#footnote-1), providing sufficient reason to deploy them. This provides enough impetus to discuss the ethical decision-making on the part of the engineer.

The societal implications of this technology involve its interaction with people. One of the key implications is how will the companies implementing this technology handle privacy? By necessity, the cars will be equipped with many sensors continuously collecting data from their surroundings to be processed. How will the companies be required to handle this data? Should companies be allowed to sell or even keep this data, especially if it includes information about uninvolved parties; e.g. pedestrians or other drivers? Another prominent implication has to do with companies’ transparency. Should companies’ entire inner workings be exposed? What about the programs that control how the car makes decisions? If machine learning is used, how much data should be provided on how it functions?

The technological implications of self-driving cars involve how effective the technology can be at meeting the design goals. The questions that arise in this field include: Can self-driving cars be designed to be sufficiently safe? What should be required of the sensors that are used? Would it be acceptable to skimp on cost for sensors and processing if it means potentially endangering people? Once a person is no longer in control of the vehicle, how does that change their liability if the car kills or injures someone? Under what circumstances should the company or the engineers be held liable?

The environmental implications of this technology involve how the technology impacts the environment at large. Concerns in this area include: What is the environmental (and economic for that matter) impact if self-driving cars follow the in the footsteps of iphones? I.e. become a pseudo-disposable technology that is replaced periodically for a newer version. Ride-sharing services may become more popular, which may provide another avenue for public transportation. Will this availability of ride-sharing combined with the ease of use of autonomous vehicles increase or decrease the carbon footprint of transportation?

The economic implications of self-driving vehicles involve how the economic market reacts to the new technology, and who stands to lose or gain money. The economic implications of self-driving cars include: Will the ride-sharing economy provide another more convenient avenue of public transportation for the working class? Should governments provide public autonomous vehicles as a form of public transportation as a way to make it affordable for the working class? If a car weighed the lives of its passengers as equal to the lives of others, (i.e. it doesn’t favor the person who bought it or paid for its service) is that a marketable product? Why would people pay for a car that would kill them over someone else?

Regarding a trolley-problem esque scenario, where a crash is inevitable and outcomes must be predicted and an ‘optimized’ solution must be reached, the main issue that arises that what is considered ‘optimized’ changes depending on who you ask. The results of the MIT study: The Moral Machine[[3]](#footnote-2) show that the mortal outcomes people most agree with vary widely based on geographical region, nationality and culture. The question arises: can, or even should, these systems be designed with everyone’s ethical considerations accounted for? The results of the Moral Machine suggests that there is not a universal ethic that can be implemented in self-driving cars to please everyone.

Rather than rely on the design engineer’s ethical opinions to hard-code in a moral code in a self-driving car, I think an approach should be taken similar to the one MIT took with the Moral Machine. Based on the location in which the cars are to be deployed, they should be programmed with a decision-making scheme that reflects the results of an appropriately sized and representative polling of the local population. The benefit of this approach is that the discrepancy between the cars’ code of ethics and the population’s code of ethics is minimized and the car would ideally simulate the reaction of the average person from that location. This is largely a utilitarian approach as it aims to minimize the potential harm that would result from an individual’s car’s ethics not aligning with their own, and making a decision they do not morally agree with.

The main problem with this solution is that it does not address the ethics of the person using the vehicle, but imposes the ethics of the average person onto the individual. This in itself is a major ethical issue as it goes against the individualist ethic on which much of our morality is based as it takes away the user’s moral agency. Other notable flaws with and criticisms of this approach include: How would these zones be decided? By whom? What happens when someone in a self-driving car changes zones? Would they have to update their car’s software to match with the local ethics?

When it comes to computer-controlled self-driving cars, one of the main issues regarding liability is that the computer is often programmed using learning algorithms for image processing rather than being directly human-programmed. There are many cases on the road where road signs or lane dividers are imperfect, and the ‘correct’ response to this scenario is difficult to discern for a learning algorithm that hasn’t been fed each specific edge-case in the source data. This is exemplified in the case in Mountain View where a Tesla drove into a concrete divider[[4]](#footnote-3). A bug caused the autopilot responded incorrectly to its environment, and crashed the vehicle. On one hand, Tesla should be responsible for sufficiently testing their product so that these situations are largely avoided, and a company should be liable for releasing a dangerous device to the public without sufficient warning of its dangerous nature. On the other hand, the edge cases that need to be accounted for are astronomically large, so catching every one of them is nearly impossible. Also, the driver had taken his hands off the wheel prior to the accident, and ignored visual and audio warnings that he was required to take action. In this case, since the driver was warned by the car to take action and he did not, I would say that the driver was liable for the accident and not Tesla.

However, this is only considering level-4 autonomous vehicles. The question of liability changes drastically when fully autonomous vehicles are considered instead. Now it is expected that there be no human interaction required for the vehicle to operate, and the question of ethical liability is determined by what the implicit agreement is between customer and company when the customer purchases an autonomous vehicle. It is not reasonable or realistic that the agreement would be “This vehicle will keep the customer completely safe under all circumstances” admonishing the customer of all responsibility. However, it is also not reasonable for the customer to assume all responsibility under the mantra “Let the buyer beware”, because an autonomous vehicle is so complex and its function is so far removed from the average consumer’s understanding, that it is impossible for the consumer to forsee the entire risk of buying one.

Since the consumer has is completely removed from the operation of the vehicle, it seems best to put the liability of the correct operation of the vehicle on the company who made it rather than on the consumer using it. Even though the company cannot promise the consumer that they will be 100% safe using the vehicle, the consumer is still placing their trust in the hands of the company to ensure their relative safety. If that trust is betrayed, the company should be held legally responsible. This should also apply to scenarios where a bystander or other vehicles are involved. The company whose car was deemed at fault for the accident should be held responsible for the accident.

1. [Path to Autonomy: Self-Driving Car Levels 0 to 5 Explained](https://www.caranddriver.com/features/path-to-autonomy-self-driving-car-levels-0-to-5-explained-feature) [↑](#footnote-ref-0)
2. [TO SAVE THE MOST LIVES, DEPLOY (IMPERFECT) SELF-DRIVING CARS ASAP](https://www.wired.com/story/self-driving-cars-rand-report/) [↑](#footnote-ref-1)
3. [The Moral Machine](http://moralmachine.mit.edu/) [↑](#footnote-ref-2)
4. [Tesla car that crashed and killed driver was running on Autopilot, firm says](https://www.theguardian.com/technology/2018/mar/31/tesla-car-crash-autopilot-mountain-view) [↑](#footnote-ref-3)