An Analysis of Morality in Autonomous Vehicles

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Michael Tang
Spring, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

An Analysis of Morality in Autonomous Vehicles

Autonomous Vehicles and the Ethical Questions They Raise

In July of 2019, Tesla claimed that they would launch full self-driving capabilities to Tesla vehicles currently on the road by the end of the year (Siddiqui, 2019). These cars contain the necessary hardware for fully autonomous driving, but require a software update in order to use that hardware. Even with the software update, however, these Teslas will still allow and require human intervention in the event of an error or emergency. The next generation of autonomous vehicles (AVs), dubbed by the U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA) as Level 5 AVs, will not require any human intervention, and thus will not need steering wheels or pedals (Lynberg, 2017). Level 5 AVs have the potential to reduce accidents by constantly monitoring their surroundings and reacting faster than human drivers can. They can also provide transport to people who are otherwise unable to drive, such as the elderly or disabled. However, a Level 5 AV would require the computer inside the car to be responsible for everything a human driver would normally be responsible for, including what happens in an accident. In a situation where injury or loss of life is inevitable, inside the car or outside, the car must make what it considers to be the most moral decision. What are possible solutions to creating an ethical AV? And when we discuss AVs, how is morality determined? Utilitarian ethics will be used to investigate these topics further.

How Can We Create a Moral Autonomous Vehicle?

The research questions that this paper addresses are as follows: What are some possible social or technological solutions to create a "moral" autonomous vehicle? When we discuss autonomous vehicles, how is morality determined? In order to answer these questions, documentary research methods are used to analyze publications about autonomous vehicle ethics

containing data from different surveys. Examples of publications containing these surveys include crash statistics from the U.S. Department of Transportation between 2005 and 2007, or information about consumer's opinion on perceived safety of AVs. Many of these publications have differing and conflicting views on the ethics of autonomous vehicles. For example, Bonnefon et al. (2016) conducted a study revolving around the trolley problem, where participants decided between AVs that chose the option that saved the most lives versus AVs that protected their passengers at all costs, while Holstein and Dodig-Crnkovic (2018) rejected the notion of the trolley problem altogether. Documentary research methods synthesize the main themes throughout all of these examples and draw conclusions for the research question in this paper. In addition, the problem of creating an ethical AV has the characteristics of a wicked problem. Although a potential solution to the research question is provided, the solution is not the only possible acceptable answer to the question. This is due to the nature of wicked problems, where there are many different possible solutions, some being better than others based on the context of the problem. Using wicked problem framing shows that the proposed solution can be the best solution based on how the question is framed using utilitarian ethics.

The Adoption of Different Levels of Autonomous Vehicles in the United States

Autonomous vehicles are not all created equal. There are several different levels of automation for autonomous vehicles, ranging from level 1 to 5. Level 1 AVs "are controlled by the driver, but some driving assist features may be included in the vehicle design," while level 5 AVs "are capable of performing all driving functions under all conditions." Level 5 AVs do not allow any sort of human intervention (Lynberg, 2017). For the purpose of this research, all levels of AVs that include some form of artificial intelligence that makes decisions for the vehicle on the road, regardless of whether the vehicle includes controls for human intervention such as

steering wheels and pedals, will be considered. This includes AVs from level 2 up to level 5. Although including this range of levels of AVs introduces problems related to liability in accidents, which will be addressed later in this research paper, it provides a more tangible solution that can be considered and implemented now or in the near future.

AVs have the potential to save many lives annually. According to the US Department of Transportation, 94% of all car crashes in the US between 2005 and 2007 were caused by driver error such as overcompensation, drunk driving, and aggressive driving. The remaining 6% of crashes were caused by problems with the vehicle itself, the environment, and other unknown reasons (Singh, 2015). Thousands of lives could be saved annually by replacing human drivers with the hardware and software in AVs.

This research paper will only be concerned with the adoption of AVs in the United States (US). This is because the US will most likely never require its citizens to purchase and use a level 5 AV, due to the capitalist nature of its industries. While the US can force car manufacturers to implement certain safety features in cars, they cannot force consumers to purchase these vehicles, no matter how much safer these features make the vehicle. For example, a backup camera in a car is a feature that reduces the blind zones of the car, and can save lives by preventing backover crashes over more vulnerable people such as children and senior citizens. The NHTSA mandates that all vehicles made after May 2018 require a backup camera ("Rearview Video Systems," n.d.). However, consumers are in no way required to purchase one of these new cars with a backup camera. In addition, if the US were to require auto manufacturers to only produce level 5 AVs, the manufacturers would most likely lobby against this decision. Manufacturers sell different trims and tiers of vehicles to people of varying levels of income, and to them, automation would most likely be adopted as an additional feature or

option. To require automation on every one of their vehicles would most likely never happen, and even if it did, consumers still have the option of not purchasing these vehicles. Therefore, this research paper will assume that the US will never require everyone to purchase and own a level 5 AV.

To summarize, this research will be concerned with level 2 to level 5 AVs. It also assumes the US government will never mandate that everyone purchase and own an AV, and that having AVs on the road will be beneficial towards lowering the number of car accidents each year.

Utilitarian Ethics for Autonomous Vehicles

The development of autonomous vehicles is rapidly accelerating for good reason. AVs can reduce the risk of car accidents by constantly monitoring their surroundings, and can provide transportation to those who are otherwise unable to drive. However, AVs would require the computer inside the car to be responsible for everything a human driver would normally be responsible for, including what happens in an inevitable accident, where the car must make what it considers to be the most ethical decision. How, then, do we create an ethical AV?

The STS framework used to analyze this question is utilitarian ethics. Fuchs and Macrina (2005) define utilitarian ethics as a consequentialist theory that "acknowledges the fact that many acts do not produce purely good consequences or purely bad consequences, but some combination of the two. To determine whether a particular act is moral, a person must sum up all the consequences, both good and bad, and assess the net outcome. Moral actions are those that cause the best balance of good versus bad consequences." Jeremy Bentham and John Stuart Mill both describe utilitarianism as "a measure of maximizing pleasure while minimizing pain" (Sandel, 2010). Bentham wrote that people prefer pleasure over pain, and that it is with this

belief that moral principles are founded (McCartney & Parent, 2015). In essence, solutions to problems framed in the context of utilitarianism ethics are neither completely right nor completely wrong. Rather, they are better or worse than other solutions. The best solution is the one that produces as much overall "good" consequence as possible. Although the ideas of utilitarian ethics sound similar to those of a cost-benefit analysis, the main difference in the context of the research question is that a cost-benefit analysis is done for the benefit of only the party performing the analysis, while utilitarianism considers everyone related to the problem. Therefore, the proposed solution to the research question is justified in the frame of utilitarian ethics. In order to be able to use utilitarian ethics to analyze the proposed solution, this research paper assumes that "good" in terms of autonomous vehicle ethics means preventing as much death and injury in car accidents as possible. Therefore, the proposed solution is centered around preventing the most death and injury overall, even if it means it will cause more harm in certain situations.

Solutions for Creating an Ethical Autonomous Vehicle

Under utilitarian ethics, where injury or loss of life is considered "bad," what is considered ethical would be to adopt as many AVs as possible. In addition, because AVs will probably never be mandated by a governing body unless they cause zero deaths, people must be willing to adopt AVs themselves in order for AVs to become widespread. Thus, in order to increase the adoption rate of AVs, we need to make the perceived safety of the vehicles higher than human driven vehicles. To accomplish this, the proposed solution to this research question is three-fold. First, AV manufacturers should institute a policy for AVs to always protect its passengers in the event of an emergency. Second, vehicle manufacturers should be transparent about the technologies and software used in autonomous vehicles. Third, there should be controls

inside AVs, even level 5 AVs, that allow a passenger to take control in the event of an emergency.

A major assumption that goes into the proposed solution is that AVs are safer than human drivers. There have been several studies done in the past to show that this is true. One such study was done by Teoh and Kidd (2014), who compared the crash rates of Google AVs with those of human drivers in Mountain View, California. They found that "the rate of police-reportable crashes per million VMT (vehicle miles traveled) for Google cars was lower than the rate of police-reported crashes of human-driven passenger vehicles traveling in Mountain View (2.19 vs 6.06), was similar to California overall (2.19 vs 1.92), and was lower than the United States (2.19 vs 3.59)." The results Teoh and Kidd found suggest that AVs are safer than human drivers, and having more AVs on the road would mean a lower crash rate overall.

Instituting a passenger-first policy for AVs will increase the trust of consumers looking to purchase an AV. Bonnefon et al. (2016) conducted six online surveys involving 1928 participants. They presented the participants with a series of different situations involving an autonomous vehicle in an inevitable crash. Participants were forced to choose between killing several pedestrians, or killing the passenger inside the vehicle. The participants "overwhelmingly expressed a moral preference for utilitarian AVs programmed to minimize the number of casualties," however, they also "indicated a significantly lower likelihood ($P < 0.001$) of buying the AV when they imagined the situation in which they and their family member would be sacrificed for the greater good." The trend of "people praise self-sacrificing AVs and welcome them on the road without actually wanting to buy one for themselves" is shown throughout the entire study. Without a passenger-first policy, is it highly unlikely that AVs will ever become widely adopted. However, instituting such a policy means that there will be situations where a

greater number of people die than if the passenger were sacrificed. For example, in a situation where an AV whose brakes have failed is careening towards a group of pedestrians on a crosswalk, the AV might make the decision of driving through the pedestrians and risk injuring or killing them rather than drive off the road or into a wall and risk killing its lone passenger. Despite this possibility, implementing a passenger-first policy would mean consumers will be much more likely to purchase AVs. By having more AVs on the road overall, the overall crash rate on the roads will be reduced due to the safer driving of AVs compared to that of human drivers, leading to less death and injury overall. Therefore, instituting a passenger-first policy is a good solution under utilitarian ethics.

The study described above shows that consumers are very unlikely to purchase AVs if they do not trust them. Implementing a passenger-first policy will increase the trust of the consumer, but there are other ways to increase trust in AVs as well. Choi and Ji (2015) conducted a study where they tried to determine what factors drive a person into trusting an autonomous vehicle. They surveyed 552 people, and asked them to agree or disagree with statements using a 7-point Likert scale. These statements included ones such as, "I believe that I can form a mental model and predict future behavior of autonomous vehicles," or "I believe that autonomous vehicles act consistently and their behavior can be forecasted." They found that "trust is a major construct for predicting the adoption of autonomous vehicles." In order to receive this trust, they also found that "it is important to provide functions that allow drivers to recover control in situations whenever they so desire," and that it is important to be very transparent about the technologies and processes used in the AV, so that the consumer has information that helps them "predict and understand the operating of the autonomous vehicles." Having controls in the car that the passengers can use, as well as having AV manufacturers be

transparent about the technology inside the vehicle, will both increase the trust of consumers in AVs, which will therefore increase the adoption rate of AVs by consumers. The more AVs there are on the road, the lower the overall crash rate, number of deaths, and number of injuries will be. Thus, keeping controls in AVs and transparency about the technology inside the vehicles are both good solutions under utilitarian ethics.

The answer to the research question on how to create an ethical AV does not solely involve technical answers. Autonomous vehicle manufacturers need to be transparent about the design and processes in the vehicle. The design of an AV will most likely involve a neural network, as autonomous vehicles need to be able to make decisions in any surroundings, on any road, in any situation. There are too many factors on the road to be able to implement a list of all the different possible scenarios that could happen to the AV while it is driving. Some of the factors include severe weather impacting the sensors on the AV, children running across the road, or other vehicles around the AV driving illegally. A neural network is a system that makes decisions based on probability. The network accepts a list of inputs from the sensors connected to it, such as the cameras or radar sensors on the vehicle, and based on all of that information, chooses a course of action that has the highest probability of "success." The decisions the vehicle can make are based on previous events the network has been trained on. In essence, the neural network makes decisions based on events it has "seen" or been trained on in the past, very similarly to how humans learn from previous events that happened to us. The more events the network is trained on, the more general of a driving policy it can form. Because neural networks act using probability based on the events they have been trained on before, it is unlikely that autonomous vehicles will ever be completely perfect. However, it is important for AV manufacturers to educate consumers on these topics, and that AVs are not perfect, but they still

do reduce the rate of accidents on the road. Constantly telling consumers that AVs are perfect and will never make any mistakes will lose the trust of the consumers. Educating consumers and maintaining transparency about the technologies inside AVs will increase the trustworthiness of AVs to consumers, leading to a higher adoption rate of AVs and a lower overall crash rate. A lower crash rate means a lower number of deaths and injuries overall. Therefore, AV manufacturers being transparent about the design and technology in their vehicles is a good solution under utilitarian ethics.

Implementing a passenger-first policy, being transparent about AV technologies, and having controls even in level 5 AVs will all improve the trust of the consumer. By doing so, consumers are more likely to adopt AVs, which means there will be more AVs on the road. A greater number of AVs driving on the road will mean a lower crash rate overall, leading to more lives saved. These three factors serve as a good solution under utilitarian ethics.

There are several constraints defined, as well as some limitations in this research paper. Instituting a passenger-first policy in AVs will mean that there may be situations where there is a greater loss of life outside of the AV than if the AV sacrificed its own passenger. Those sorts of situations are very similar to the one described by the trolley problem. Originally designed by Philippa Foot, a professor of philosophy at the Massachusetts Institute of Technology, the trolley problem positions the reader as the driver of a trolley. The trolley is approaching five workmen on the track very quickly, and there is no way to stop the trolley. The reader has the option of pulling a lever to divert the trolley onto a side track, thereby saving the five men from death. However, there is a single worker on the side track who will be killed if the reader diverts the trolley. The problem forces the reader to either take inaction and let the five workmen die, or pull the lever and kill the one man on the side track (Thomson, 1985). In terms of AVs, the AV is the

"reader" in the trolley problem, the pedestrians in front of the AV are the five workmen, and the passenger inside the AV is the one workman on the side of the track. Although there may be situations where more people die due to the passenger-first policy, it would be better overall due to the higher adoption rate of AVs.

It is important to note that the trolley problem scenarios described above are inherently unfair. No matter which decision is made, the trolley problem always ends in failure because one group of people died. If the AV follows its passenger-first policy and kills a group of pedestrians, the fact still remains that the pedestrians died. It is important to discuss the ethics of AVs outside of these trolley problem scenarios so that these scenarios can be prevented from happening in the first place. Mirnig and Meschtscherjakov (2019) argue that "confronting automated vehicle technology with the trolley problem and then judging said technology on their decision taken within the dilemma is akin to posing the same problem to an accused human individual at court, and then convicting them anyway regardless of their response because there was never a right choice to begin with." They mention that the trolley problem should be used "not as a collection of two possible outcomes but rather as one possible outcome among several. The trolley problem is the one that needs to be avoided." Future studies surrounding AV ethics should investigate actions an AV should take in order to avoid ending up in a dilemma situation. Part of these actions includes researching the possibility of allowing AVs to purposely drive aggressively or illegally to avoid accidents. For example, an AV could speed up and go over the speed limit instead of slowing down to avoid a car approaching from the side.

Another limitation of this research paper is that it only considers AV adoption in the United States, due to the low likelihood of the US creating legislation requiring its citizens to purchase an AV. This was to support the use of utilitarian ethics in this research paper, so that

the only reason why there would be an increase in AV adoption rate is because people made the choice for themselves to purchase an AV. Future studies should look into regulations and requirements for AVs by governments outside of the US, and how that might affect the ethical policies of AVs.

There is also the question of liability in an accident arising from the solution proposed by this research paper. Having controls in the car that the passengers can use at any time will improve the trust of the passengers in the AV, but also raises the question of who is at fault in a crash. Schroll (2014) investigates how current insurance and liability policies would apply to AVs. In this paper, he discusses how in an autonomous vehicle with controls, the "driver" in an AV is unknown, and how all riders in the AV should potentially be held for responsibility. Liability with regards to AVs is out of the scope of this research paper, but is definitely an important topic that needs to be addressed in the future.

The Future of Autonomous Vehicles

The proposed solution on how to create an ethical AV involves implementing a passenger-first policy in AVs, vehicle manufacturers being transparent about the technologies used in and the benefits of AVs, and there being controls inside all AVs that allow a passenger to take control at any time in the event of an emergency. AVs are able to provide transportation to the disabled or elderly, and have already been proven to have a lower crash rate than human drivers. While the technical challenges of AVs are being resolved through constant research and innovation, the social challenges AVs face still stand. Implementing a solution such as the one proposed in this paper can increase peoples trust in and adoption rate of AVs. This paper does not address every issue surrounding AV ethics, but it does serve as a step in the right direction towards saving lives.

Works Cited

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. Science, 352(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Choi, J. K., & Ji, Y. G. (2015). Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *International Journal of Human–Computer Interaction*, *31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Fuchs, B. A., & Macrina, F. L. (2005). Ethics and the scientist. *Scientific Integrity: Text and Cases in Responsible Conduct of Research*, 19–38.

Holstein, T., & Dodig-Crnkovic, G. (2018). Avoiding the Intrinsic Unfairness of the Trolley Problem. Proceedings of the International Workshop on Software Fairness, 32–37. https://doi.org/10.1145/3194770.3194772

Lynberg, M. (2017, September 7). Automated Vehicles for Safety [Text]. Retrieved December 10, 2019, from NHTSA website: https://www.nhtsa.gov/technology-innovation/automated-vehicles

McCartney, S., & Parent, R. (2015). 2.2 Utilitarian Ethics. In *Ethics in Law Enforcement*. Retrieved from https://opentextbc.ca/ethicsinlawenforcement/chapter/2-2-utilitarian-ethics/

Mirnig, A. G., & Meschtscherjakov, A. (2019). Trolled by the Trolley Problem: On What Matters for Ethical Decision Making in Automated Vehicles. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 509:1–509:10. https://doi.org/10.1145/3290605.3300739

Rearview Video Systems. (n.d.). Retrieved January 31, 2020, from https://www.safercar.gov/Vehicle-Shoppers/Safety-Technology/rvs

Sandel, M. J. (2010). *Justice: What's the right thing to do?* Macmillan.

Schroll, C. (2014). Splitting the Bill: Creating a National Car Insurance Fund to Pay for Accidents in Autonomous Vehicles Notes and Comments. *Northwestern University Law Review*, (3), 803–834.

Siddiqui, F. (2019, July 17). Tesla floats fully self-driving cars as soon as this year. Many are worried about what that will unleash. Retrieved December 10, 2019, from Washington Post website: https://www.washingtonpost.com/technology/2019/07/17/tesla-floats-fully-self-driving-cars-soon-this-year-many-are-worried-about-what-that-will-unleash/

Singh, S. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey*. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

Teoh, E. R., & Kidd, D. G. (2017). Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research*, *63*, 57–60. https://doi.org/10.1016/j.jsr.2017.08.008

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.