A Utilitarian Ethics Analysis of The First Fatal Autonomous Vehicle Accident

STS Research Paper
Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia

By

Samuel Longo

February 18, 2021

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.

Signed: Samuel Longo

Approved: _____ Date _____
Benjamin J. Laugelli, Assistant Professor, Department of Engineering and Society

Introduction

In the past decade, vehicles with automated electronic safety features have become increasingly common. Automobile manufacturers, university research groups, and global militaries have conducted research on lowering the risk to vehicle operators through improved design and new technology while aiming to ultimately eliminate the need for a driver. In 2018 the first fatality involving an autonomous car came into national attention when one of Uber's test vehicles operated by Rafaela Vasquez struck bicyclist Eleaine Herzberg in Tempe, Arizona (Conger, 2020). Although the vehicle was self-driving and requiring an operator as the technology was not ready for market, the accident occurred because the driver was not paying attention and was taking a risk she would not normally take in a standard automobile: watching Hulu on her phone while driving. In recent years, scholars and researchers have examined the risk preferences of autonomous vehicle operators and have recognized that drivers are more willing to make risky actions, which can be harmful to others while driving autonomous or partially autonomous vehicles. Scholars have also examined the implications this risk seeking behavior will have on pedestrians and other automobile operators. However, this event and those involved have yet to be examined with questions of efficiency and morality in the long term and short term of actor decisions.

In this paper, I will examine the case of Elaine Herzberg's death through the lens of utilitarianism to describe the morality of Rafaela Vasquez's actions as well as the actions of Uber and Volvo. I will argue that Vasquez acted immorally in the eyes of utilitarianism while Uber and Volvo were justified in their actions according to utilitarian theory. I will do this by examining the long term and short-term efficiency of decisions made by actors for both themselves and society at large. I will also examine the case through a lens of Pareto efficiency also known as the no harm principle of utilitarianism and the lens of Kaldor-Hicks

efficiency, which is a branch of utilitarianism with no regards for utility distribution (van de Poel and Royakkers, 2011). The consequences of this paper will be to open the discussion around autonomous vehicle ethics to include the analysis of tangible case studies rather than mainly generalized studies about what consumers will be willing to purchase. This paper will help to inform policy makers about the incentive structures and societal welfare structures around the autonomous vehicle market.

Background

On March 18 of 2018, Uber supervising driver Rafaela Vasquez was operating a Volvo XC90 SUV, which had self-driving technology that was being tested by Uber, the E-taxi service known for autonomous vehicle research and its matching system of part-time drivers to riders. Vasquez was watching an episode of *The Voice* on Hulu through her phone with both hands off the steering wheel of the vehicle when pedestrian Elaine Herzburg was fatally struck crossing the road while walking her bike across the street. Vasquez was tasked as a supervising driver for the partially automated vehicle as it was an experimental technology Uber was testing and Vasquez was not paying attention to the road when the accident occurred. Herzburg was crossing a darkened stretch of road and was not in a crosswalk and the Volvo's computer system did not register Herzburg as a pedestrian. (Wilson et al, 2020)

Literature Review

Multiple scholars have discussed utilitarianism in the realm of autonomous vehicles. These discussions typically involve the classic trolley car problem where scholars and survey makers are pitting utilitarian ideals of safety in autonomous vehicles against self-safety ideals of autonomous vehicles. Self-safety ideals in autonomous vehicles would entail that the vehicle

values the lives of its own passengers above all others and is thus in conflict with utilitarianism. The following analyses discuss how vehicles exercising utilitarian ethics are likely to be societally accepted and are unlikely to be purchased. These analyses also explain the long-term utility of having autonomous vehicles as a norm on roads. These analyses examine the autonomous vehicle market as a whole with a utilitarian framework but fail to discuss the actions of individual actors and the utilities those actions produce.

Through their academic research titled The Social Dilemma of Autonomous Vehicles, MIT researchers Bonnefon, Shariff, and Rahwan conducted a set of surveys with over 1000 participants exploring the preferences of potential autonomous vehicle consumers for different types of ethics (Bonnefon et al, 2016). While consumers did not come to any unanimous agreements, Bonnefon et al determined that 76% of participants viewed a self-driving vehicle that sacrificed the driver for the lives of 10 pedestrians as moral, thereby expressing a Kaldor-Hicks utilitarian view. However, these same participants are less willing to purchase a utilitarianly ethical vehicle compared with one that protects the driver as shown in the study quote, "people praise utilitarian, self-sacrificing AVs and welcome them on the road, without actually wanting to buy one for themselves." In discussion of the MIT study, author Evan Ackerman writes "We want autonomous cars to be as safe as possible, as long as they're safest for us first." (Ackerman, 2016). Ackerman validates consumer bias for self-safety. When discussing the trolley car scenarios used in the MIT surveys, Ackerman makes the conjecture that autonomous cars need to "make decisions that maximize safety in a way that is understandable." Ackerman argues that the utilitarian decision making in autonomous vehicles does not need to be absolute, but can be on a spectrum while still having a preference for the driver's protection. Ackerman also argues that if utilitarianism is the ethical framework of the artificial intelligence, that self-driving vehicles will have low demand and will not be brought to

market as quickly, which would ultimately cost lives assuming that these vehicles can drive safer than humans can on average. The MIT study also contributes that consumers do not want government regulation dictating utilitarian ethics for autonomous vehicles and that if such regulation came into existence, consumers would be less likely to purchase a self-driving vehicle (Bonnefon et al, 2016).

In his paper titled Will My Next Car be a Libertarian or a Utilitarian: Who will decide, author Tom Fournier explains that autonomous cars are in line with utilitarian ideology through their reduced environmental impact and reduced total cost for accidents compared with standard cars (Fournier, 2016). Fournier explains that the National Highway Traffic and Safety Administration estimates that the total cost for traffic related accidents was $277 billion in 2010 alone. Fournier also argues that our endowment allocation as a society for autonomous vehicles is not consistent with utilitarian values given that the war on terror in response to 3,380 lost American lives cost about $95 billion dollars in 2014 alone. This argument implies that subsidizing the autonomous vehicle market would be a more just action than military spending given the annual cost both nominated in dollars and in human life of automobile accidents. Fournier goes further down the unique path that autonomous vehicles are more efficient for the environment and should reduce pollution from taxi services and automobile ownership, predicting the possibility that in the future individuals will not own cars and will use on-demand fleets of driverless taxis (Fournier, 2016). This would theoretically reduce emissions, and the demand for resources to make cars such as rare metals. These positive environmental impacts are utilitarian in that they add net utility spread over most of society. There are also the mentioned costs and hassles associated with automobile ownership that would be reduced through this utilitarian on-demand fleet according to Fournier.

The first source authored by Bonnefon et. al illustrates that there is a dilemma of

preference among consumers for utilitarian vehicles for everyone else, and self-protecting cars for the individual consumer. Fournier argues that autonomous vehicles bring quantifiable utility into the market through environmental and economic effects. While it is necessary to describe the large-scale ramifications of utilitarian software for autonomous vehicles and the effects on utility that autonomous vehicles will have on society in general, I will deploy the utilitarian ethical framework to examine the morality of Vasquez, Uber, and Volvo in their decisions leading up to the fatality of Elaine Herzburg. As the Bonnefon et. al 2016 MIT research group has said, "we are about to endow millions of vehicles with autonomy, a serious consideration of algorithmic morality has never been more urgent… these types of decisions need be made well before AVs become a global commodity" (Bonnefon et al, 2016). The case study style analysis of this paper will contribute to the discussion of what level of utilitarian ethics should be employed in the market of autonomous vehicles and will guide engineers making software and business decisions that will attempt to place fair values on human life. This analysis will fill the gap left by prior literature through its approach of examining individual actors and the utilities they create.

Conceptual Framework

The actions of Vasquez, Uber, and Volvo will be addressed by examining the short-term and long-term efficiencies of the actions as well as the Pareto and Kaldor-Hicks efficiencies of the actions. The long-term and short-term efficiencies in the overall system/society will also be discussed. The ethical framework of utilitarian ethics will be employed to discuss the morality of actors involved in the Tempe, Arizona collision.

The utilitarian ethical framework, first founded by Jeremy Bentham, focuses on the

usefulness of actions and the pleasure that actions bring (van de Poel & Royakkers, 2011). In this situation, the actions under examination are the actions of Vasquez who chose to watch videos on her phone rather than operate the vehicle, Uber who chose to test this self-driving technology on the streets of Tempe provided that a supervising driver was behind the wheel in the event of malfunction, and Volvo who created the self-driving technology and allowed that technology to be put on the market for live real-world testing. The utilitarian framework derives from hedonism and directs that people should take the most pleasure bringing actions on the grounds that the greatest pleasure is given to the greatest number of people. Within utilitarianism there is also the concept of John Stuart Mill's no harm principle, which means that a person may take any pleasure bringing action so long as no harm is done to others (van de Poel & Royakkers, 2011). The no harm principle is a foundation of Pareto efficiency whereby the best outcome maximizes efficiency while causing no harm to others (Valentin, n.d.).

There is also Kaldor-Hicks utilitarianism where the criteria for morality lies only in how much pleasure is brought to how many people and a decision may still be righteous if someone gets hurt but another gains more pleasure than pain is given to others. This sect of utilitarianism discusses the moral balance sheet where the costs and benefits of actions can be weighed against one another to choose the decision yielding maximum utility. Some categories of the moral balance sheet that I will use include harm to individuals and entities like Elaine Herzburg or Uber, benefits to American automobile consumers, and degrees of pleasure from actions. In Kaldor-Hicks utilitarianism, the most righteous decision is the one that maximizes utility in a full system but ignores the value of distributive justice (Valentin, n.d.). A decision or system can be Kaldor-Hicks efficient even with only one beneficiary gaining a large amount of utility while many others have a small pain as a result.

Decisions may be both Pareto and Kaldor-Hicks efficient so long as no one is hurt and

maximum utility is reached. Decisions may also be neither, where the decision does not maximize utility subject to no harm principle or not. Decisions may also be pareto efficient only or Kaldor-Hicks efficient only. Long-term efficient decisions and short-term efficient decisions can also be in conflict with one another under the utilitarian frame (Valentin, n.d). Because of this, it can be concluded that individuals may be in line with certain utilitarian schools while out of line with others or may be completely unethical in the frame of utilitarianism.

Through the lens of utilitarian ethics, I will examine how the decisions made by Vasquez, Uber, and Volvo can be considered just, in accordance with the different utilitarian schools discussed thus far. I will accomplish this by taking the key actions made by each individual and confirming these actions as in line or out of line with the discussed utilitarian theories. Ultimately, I will use utilitarian ethics to determine which actors are morally just and within what constraints.

Analysis

Tempe Arizona saw early testing of autonomous vehicles in live real-world settings. These tests proved fatal for bicyclist Elaine Herzburg because of supervising driver Rafaela Vasquez's nonutilitarian decisions, which violated the no harm principle and caused a negative net utility for autonomous vehicle producers, consumers, and society as a whole. By committing actions that were both Pareto and Kaldor-Hicks inefficient, Vasquez violated utilitarian ethics. In contrast, the other key decision makers, Uber and Volvo, both made decisions in line with utilitarian theory. The decisions of Uber and Volvo were short-term, long-term, Pareto, and Kaldor-Hicks efficient. According to van de Poel and Royakkers, a lack of utility and a loss inflicted on another is a violation of both Bentham's and Mills's utilitarian theories (van de Poel & Royakkers, 2011). Vasquez caused a lack of net utility and a loss of life for another person.

Through the lens of Utilitarianism, Vasquez is therefore morally irresponsible. The following paragraphs will consider combinations of long or short term efficiency and Pareto or Kaldor-Hicks efficiency. This analysis will examine the actions of Vasquez, Uber, and Volvo to argue that Vasquez acted immorally and Uber and Volvo did not.

Short-Term Pareto Efficiency

Short-term Pareto Efficiency is a subsect of utilitarian ethics based on maximizing utility subject to the no harm principle and constrained to immediate or near immediate effects of decisions (Valentin, n.d). Rafaela Vasquez's decisions are nonutilitarian under this ethical subset, while Uber and Volvo are utilitarian in this case.

Following the fatal accident resulting in Elaine Herzburg's death, it was determined that Vasquez had no hands on the wheel of the vehicle at the time of collision and was not watching the road as supervising driver (Conger, 2020). Vasquez's work relationship with Uber required Vasquez to be able to operate the vehicle in the event of a software failure in the car. Vasquez's choice to be driving with distraction was not pareto efficient in the short term because this choice resulted in the death of Elaine Herzburg, which violates the no harm principle by forcing a negative utility onto Herzburg and her loved ones. Also, not only do Vasquez's actions produce a negative utility for Herzburg, but a negative utility overall in the system spread further between, Vasquez herself, Uber, and Volvo. The negative utility onto Vasquez is responsibility for a death, loss of a job, and the legal consequences. The negative short-term utility onto Uber and Volvo is that they were involved in this incident, which got major national attention and damaged the brands (Wilson et al, 2020). According to the National Highway Traffic and Safety Administration, in 2019, 8.5% of motor vehicle crashes involved distracted driving (Covington, 2021). It is a reasonable assumption that Vasquez would know

that distracted driving would put others at risk and would thus be pareto inefficient.

Uber and Volvo however, made decisions that were short term pareto efficient. Uber and Volvo recognized that the model of autonomous vehicle being tested was imperfect and could have dangerous software flaws (Conger 2020). Uber thus hired a supervising driver to protect against the possibility of an accident. The actions of Uber and Volvo were therefore in line with the no harm principle in that they would protect against causing accidents by not relying solely on the vehicle. Uber and Volvo also contributed to short term efficiency in the market by providing consumers in Tempe a new choice for transportation that was potentially safer than some other existing methods. Having this new choice for a transportation method should allow individuals to better maximize their utility by choosing the transportation method most optimal for themselves.

Long-Term Pareto Efficiency

Similar to short-term pareto efficiency, Long-term Pareto Efficiency is a subsect of utilitarian ethics based on maximizing utility subject to the no harm principle, but long-term efficiency is not subject to the immediate time constraints on effects of decisions and for the purpose of this section will be in the range of around 50 years (Valentin, n.d). Rafaela Vasquez's decisions are still nonutilitarian here, while Uber and Volvo are utilitarian.

Vasquez's actions are instantly disqualified from being long-term pareto efficient because they still cause harm and are inefficient in the short-term. In the long-term, Vasquez is still responsible for the death of Herzburg and contributed to net negative utility in the system.

Uber and Volvo are in appropriate standing with utilitarianism in this case as both are still unresponsible for causing harm and are bringing utility to many consumers and the overall

market. Uber and Volvo remain unresponsible for the death of Herzburg and at this point have contributed to the human knowledge around autonomous vehicles and thus added great net utility to the market. In December of 2020, Uber announced plans to sell its autonomous vehicle project to Aurora, a young self-driving-technology developer (Marshall, 2020). Aurora cited Uber's safety systems that were partially developed through the Tempe testing as a key part of the acquisition (Marshall 2020). As of May 2020, Volvo plans to release the 2022 XC90 (the same model tested in Tempe) with full self-driving capabilities (Balwin, n.d). It is likely that Volvo will provide utility to consumers in the future through the production of self-driving cars given that a 2019 survey found that 57% of Americans would be willing to ride in an autonomous vehicle, a growing number from prior years (Favre, 2021).

Short-Term Kaldor-Hicks Efficiency

Short-term Kaldor-Hicks efficiency is a subsect of utilitarian ethics based only on maximizing utility in a system or market, constrained to immediate or near immediate effects of decisions (Valentin, n.d). Yet again, Vasquez's decisions are nonutilitarian under this ethical subset, while Uber and Volvo are utilitarianly moral.

As previously defined, Vasquez's actions yield a negative net utility for the whole system, which could have been avoided if Vasquez as an operator had stayed focused on driving. The only positive utility within the system was the utility gained by Vasquez from watching a video on her phone. This pleasure was of little intensity or duration, which according to Bentham is therefore of little utility (van de Poel & Royakkers, 2011). Vasquez also caused great harm to Uber, which was forced to go under investigation by the National Transportation Safety Board and pause testing for multiple months while reevaluating the safety of their systems (Marshall, n.d.). Uber's brand was also immediately damaged with a vehicle of their fleet being involved in

the first fatal autonomous vehicle accident.

In the short term, Uber and Volvo satisfy Kaldor-Hicks efficiency through their decisions for the same reasons discussed in the short term pareto efficiency section. It is also worth noting that Uber and Volvo through their testing employed various peoples through their programs as software engineers, data analysts, manufacturers, est. As of December 2020, Uber had employed around 1,200 individuals for its self-driving project alone (Marshall, n.d.).

Long-Term Kaldor-Hicks Efficiency

Long-term Kaldor-Hicks efficiency is equivalent to short-term Kaldor-Hicks efficiency, with the exception of being subject to time constraints. For the purposes of this section, long-term will mean around 50 years. Uniquely, in this situation Vasquez's net effect is only slightly negative or negligible and Uber and Volvo are still positive contributors to net utility.

In the long-term Vasquez's actions are likely negligible for societal total utility. As a single actor in a menial role, there was not much authority she could have in the long term, especially when compared with goliaths like Uber and Volvo. However, this incident and Vasquez's decisions will likely contribute to regulations on the autonomous vehicle market. This incident, with Vasquez largely at fault, is partially credited with moving Uber to sell the self-driving division (Marshall, 2021).

Uber and Volvo in the long-term contribute to Kaldor-Hicks efficiency for the same reasons as listed in prior sections. Essentially, they have contributed to knowledge and exchanges in the growing autonomous vehicle market.

It could be argued that Uber does not contribute to Kaldor-Hicks efficiency in the long-term given that Uber has sold their self-driving car division. It seems fair to say that Uber will not be contributing more to the autonomous vehicle industry at first glance, but they have contributed

much utility in the long-term. Uber assembled a team of 40 expert roboticist from the National Robotics Engineering Center at Carnegie Mellon and one of them (Drew Bagnell) went on to be the CTO of Aurora, which is an early industry leader in self-driving technology today. Uber also generated a mass of early knowledge that will educate Aurora going forward and is investing 400 million USD into Aurora as part of the sale (Marshall, 2021). Uber has certainly made a major positive impact on the future of self-driving cars across the globe.

Conclusion

Through a utilitarian ethics framework, I have argued the immorality of supervising driver Rafaela Vasquez as well the morality of Uber and Volvo in the case study of the first fatal autonomous vehicle accident in the US. Through poor decision making, Vasquez created a pattern of negative utilities that harmed not only herself but others as well. Through proper planning and hard work, Uber and Volvo have helped form positive net utilities projected onto the market while preventing harm. Examples of actions from Vasquez, Uber, and Volvo have educated this argument and proven that utilitarian morality in this case study can only be claimed by Uber and Volvo.

It is important for engineers to understand that in transitions to new technologies, proper planning is needed and the technologies that engineers create will have immediate and lasting impacts on markets and society. It is important for engineers to also understand that nonutilitarian decision making can result in damage to the market as a whole and can set back the major potential gains for society that a technology may produce. It is necessary for engineers to form utility creating technologies such as autonomous vehicles, which present utility to the environment, drivers, and businesses. While it is a challenge to reflect on a concept like utility, utilitarian ethics can help one to maximize value in a society while minimizing harm. Through

utilitarian ethics, engineers can act justly in their pursuit of creating utility yielding technologies for society.

Word Count: 3764

References

Ackerman, E. (n.d.). People Want Driverless Cars with Utilitarian Ethics, Unless They're a Passenger. Retrieved March 19, 2021, from https://spectrum.ieee.org/cars-that-think/transportation/self-driving/people-want-driverless-cars-with-utilitarian-ethics-unless-theyre-a-passenger

Baldwin, R. (2020, November 13). 2022 Volvo Xc90 will get autonomous Highway PILOT hardware with lidar. Retrieved March 19, 2021, from https://www.caranddriver.com/news/a32381082/volvo-xc90-autonomous-highway-pilot-future/#:~:text=2022%20Volvo%20XC90%20Will%20Get%20Autonomous%20Highway%20Pilot%20Hardware%20with%20Lidar&text=The%20lidar%20sensor%20will%20give,over%2Dthe%2Dair%20update.

Bonnefon, J., Shariff, A., & Rahwan, I. (2016, June). *The social Dilemma of autonomous vehicles* [Scholarly project]. Retrieved March 1, 2021.

Conger, K. (2020, September 15). Driver charged in uber's fatal 2018 autonomous car crash. Retrieved March 19, 2021, from https://www.nytimes.com/2020/09/15/technology/uber-autonomous-crash-driver-charged.html

Covington, T. (n.d.). Distracted driving statistics + Research [updated 2021]: The zebra. Retrieved March 19, 2021, from https://www.thezebra.com/resources/research/distracted-driving-statistics/#:~:text=An%20estimated%20391%2C000%20drivers%20were,of%20fatal%20motor%20vehicle%20crashes.

Favre, L. (n.d.). Most Americans think self-driving cars will Be 'Commonplace' within the next decade. Retrieved March 19, 2021, from https://www.usnews.com/news/cities/articles/2019-06-06/poll-finds-americans-are-divided-on-autonomous-cars

Fournier, T. (2016). Will my next car be a libertarian or a Utilitarian?: Who Will decide? Retrieved March 19, 2021, from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7484885&casa_token=qcIDPm4mgusAAAAA%3A_dRN901Mmj7T-hocByC8wtZv-2-2ZQJzV_FI2bsxrVHsU5DjNl5_ue-MkKX9zYp_oZVZnGtByQs&tag=1

Marshall, A. (n.d.). Uber gives up on the self-driving dream. Retrieved March 19, 2021, from https://www.wired.com/story/uber-gives-up-self-driving-dream/

Valentin, S. (n.d.). The difference between the Pareto and the kaldor-hicks standard of efficiency. Retrieved March 19, 2021, from https://www.grin.com/document/491913

van de Poel, I., & Royakkers, L. (2011). *Ethics, technology, and engineering: An Introduction*. Blackwell Publishing.

Wilson, K., Schmitt, A., & Short, A. (2020, September 17). Human driver Of 'Driverless' car charged in 2018 ped death. Retrieved March 19, 2021, from https://usa.streetsblog.org/2020/09/16/human-driver-of-driverless-car-charged-in-2018-ped-death/#:~:text=Elaine%20Herzberg%2C%20who%20was%20killed,the%20time%20of%20the%20crash.&text=At%20the%20time%20of%20the%20impact%2C%20Herzberg%20was%20attempting%20to,bicycle%20rather%20than%20riding%20it.