Using Virtue Ethics to Examine the Uber's Fatal Self-Driving Crash

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

By

Daniel Choi

April 11, 2019

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: _____

Approved: _____ Date _____

Benjamin J. Laugelli, Assistant Professor, Department of Engineering and Society

Introduction

On the evening of March 18, 2018, a 49-year-old Elaine Herzberg was walking a bicycle across a four-lane road in Tempe, Arizona, when a self-driving Volvo XC90 operated by Uber ATG, Inc. fatally struck her (NTSB, 2018). Even while the vehicle was operating with an emergency backup human driver in place, the autonomous vehicle did not slow down or attempt to swerve to avoid Elaine. Her death marked the first recorded pedestrian death by a self-driving car.

In the aftermath of the accident, several revelations came to light. Firstly, it was revealed that the Arizona Governor, Doug Ducey, allowed Uber to operate its fleet of self-driving cars in Arizona in secret for several months before a public announcement was made. Furthermore, it was revealed that Uber's self-driving cars were struggling to meet the technical challenges of automation at the time of the crash (Wakabayashi, 2018). Comparing Uber's automated driving system (ADS) to other self-driving projects like Google's Waymo, Uber's self-driving cars required much more frequent human intervention. While Waymo's ADS only required human intervention every 5600 miles on average, Uber's ADS struggled to meet its target of 13 miles for each human intervention (Wakabayashi, 2018). It was also revealed that Uber operated its self-driving program with significant flaws in their testing system and protocol; Uber used poor practices when hiring and training its safety drivers, used only one safety driver per vehicle over using two, disabled automatic emergency braking, had a generally poor safety culture at Uber, delayed urgent situation alarm by 1 second, and had bad path predictions for detected cyclists (Templeton, 2019). Amongst the vast number of technical issues, one issue was particularly critical to the cause in the fatal collision with Elaine Herzberg: Uber ADS failed to account for the possibilities of pedestrians jaywalking at roads with signs prohibiting pedestrian crossing. In

a dashboard camera footage released by Tempe Police, it was also revealed that the safety driver was looking down, and away from the road, not being paying attention to the road ahead (NTSB, 2019), and also it was revealed that safety driver, Rafaela Vasquez, was a former felon who had served four years in Arizona prison for attempted armed robbery (Lin, 2018).

In the aftermath of the crash, Uber temporarily suspended its self-driving road test operations in Tempe, Arizona, as well as Pittsburgh, Toronto, and San Francisco, the four cities where Uber ATG operated its self-driving cars (Said, 2018). An investigation by the National Transportation Safety Board (NTSB) was also opened to investigate the crash, and to discuss possible recommendations to legislatures for improvement in safety of self-driving tests. Soon after the collision, Uber reached a settlement with the victim's family, and was later found to be not criminally liable for her death in court.

Through examination of Uber's Fatal Crash through the lens of care ethics framework, I will show that Uber can be held more morally responsible for causing the fatal collision that caused the death of Elaine Herzberg, instead of the individual safety driver that was operating the vehicle at the time of the accident. Within care ethics, I will explore how Uber failed at the four different stages of care ethics: Attentiveness, responsibility, competence, responsiveness. To connect the events that transpired to the ethical framework, I will use Actor Network Theory to show that the accident was not caused by a singular point of failure in an individual (carelessness of the safety driver), but rather from a collection of rogue actors within Uber, as well as a number of outside actors.

Background

Uber ATG, or Uber Advanced Technologies Group, is a subsidiary of the ride sharing company, Uber. Uber ATG is currently focused developing self-driving cars with self-driving

cars currently operating with a safety driver in Pittsburgh, Toronto, and San Francisco (Ericavella, 2018). Uber ATG is currently competing amongst a number of other companies developing self-driving cars including Google's Waymo, General Motors' Cruise, and Tesla's Autopilot. Self-driving platforms rely on various arrays of sensors like cameras and LIDAR to collect a data stream about the outside conditions, which are fed into an artificial intelligence (AI) computing unit that interprets the situation and provides an appropriate command for the car to execute. By the fundamental nature of how AI improves iteratively through "training" on their given data, the self-driving AI improves through practicing on the road to simulate possible real life conditions. As a result, Uber ATG has driven two million autonomous miles across its autonomous fleet in their effort to train their AI.

Literature Review

In the aftermath of the accident, an abundance of literature and reports have been written to analyze the factors that ultimately led to the death of Eliza Herzer. Some analysis focused on the technical and programmatic errors that led to the incident, while some analysis focused on the ethical frame work for assigning responsibilities related to accidents involving self-driving cars.

In *Collison Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian*, a report released by NTSB, the board provides a breath of clarity to the case by laying out the relevant details regarding the case, as well as their interpretation on the probable cause of this crash, and identifying several safety issues internal and external to Uber that ultimately led to the accident. Some of the relevant findings that the board found were: "Emergency response to the crash was timely and adequate", "Pedestrian's unsafe behavior in crossing the street in front of the approaching vehicle at night—violated Arizona statues was

poassibly due to diminished perception and judgement resulting from drug use", Uber ATG "did not adequately manage the anticipated safety risk of its automated driving systems' functional limitations", "Vehicle operator's prolonged visual distraction, led to her failure to detect the pedestrian in time to avoid the collision", and "Arizona's shortcomings in improving safety of ADS testing and safeguarding the public"

From their findings, the NTSB were ultimately able to conclude that the "probable cause of the crash in Tempe, Arizona, was the failure of the vehicle operator to monitor the driving environment and the operation of the automated driving system because she was visually distracted throughout the trip by her personal cell phone", but also citing multiple other contributing factors including "Uber Advanced Technologies Group's (1) inadequate safety assessment procedures, (2) ineffective oversight of vehicle operators, and (3) lack of adequate mechanisms for addressing operators' automation complacency—all a consequence of its inadequate safety culture" as well as outside contributing factors such as "(1) the impaired pedestrian's crossing of N. Mill Avenue outside a crosswalk, and (2) the Arizona Department of Transportation's insufficient oversight of automated vehicle testing."

In *Who's at Fault in Uber's Fatal Collision?*, Lin lays out the ethical case for each possible party's involvement in the accident. Like the NTSB report, Lin also recognizes that Uber and the safety driver had majority share for the blame for this incident, but does this through an ethical analysis approach by analyzing whether each party fulfilled their ethical responsibilities. The author lays out the expected societal responsibilities for each party, and analyzes whether the impact of their mistakes were critical in causing the accident.

While both literature generally agrees that majority of the fault lied with the individual safety driver's lack of attentiveness, the paper will use actor network theory to also connect other

outside factors such as Uber's safety culture, Arizona State Government, Eliza Herzberg, Volvo, and National Highway Traffic Safety Administration to the cause of the accident.

Conceptual Framework

The morality of each involved party can be analyzed using the theory of care ethics since each party had an expectation of care that they were responsible for as a societal expectation agreed upon as a society. Specifically to the Uber's fatal crash, the self-driving test was a large scale engineering project with great complexity that required a network of multiple large interconnected parties, each with different relationships and expectations to make the project function effectively and safely. Therefore, with care ethics, each involved parties' responsibility of care can be analyzed, and it can be determine whether they fulfilled their expected responsibility.

Care ethics as described by Carol Giligan and Nel Noddings emphasizes an ethical framework that emphasizes relationships, interconnectivity, specific people and contexts, and practicing empathy. The main tenants of care ethics describes a duty of care to parties that we have relationships with, and that duty depends on the context of the relationship (van de Poel & Royakkers, 2011). More specifically, I will concentrate on whether each identified parties reached the different levels of care identified by Tronto. The four different levels of care are: attentiveness, responsibility, competence, and responsiveness (Sander-Staudt, n.d).

For attentiveness, the identified party will be evaluated upon whether they became aware of a need, based on the context of the relationship and what domain of needs they were reasonably expected to have a knowledge of. For responsibility, the identified party will be evaluated beyond becoming aware of a need, and be evaluated upon whether they were willing to take care of a need that they have identified. For competence, the identified party will be

evaluated on whether they had the technical knowledge to successfully provide care for a need that they are aware exists, and were willing to provide the care. For the Tronto's last stage of care, responsiveness is the consideration the position of other actors within the same space, and "recognizing a potential for abuse in care." (Sander-Staudt, n.d).

Another conceptual framework that will be utilized in the analysis will be Actor Network Theory (ANT). ANT describes a network as a collection of human and non-human actors that are each have a different level of power, which depends on the strength of interconnections amongst the actors.

Analysis

To apply the care ethic framework, I will first use Actor Network theory to identify each of the major actors that were relevant to the self-driving project network, then apply Tronto's definition for the four different stages of care, and whether each of the identified actors fulfilled their expected levels of care. Specifically, I will attempt uncover whether the individual actor, Rafaela Vasquez, the safety driver operating the vehicle at the time of the accident, had a larger contribution to the accident, or if Uber's incompetence led to a situation that was inevitable through the scope of care ethics framework.

The first major actor identified within the network to be analyzed is the individual actor Rafaela Vasquez, the safety driver sitting at the wheel at the time of the crash. As identified by the NTSB report, several factors involving the individual actor can be ruled out of relevance in causation of the accident: "(1) driver licensing, experience, or knowledge of the automated driving system operation; (2) vehicle operator substance impairment or fatigue." According to NTSB reports, as well as the released inward-dashboard facing camera footages, it was shown

that Rafaela Vasquez was looking downwards on her cell phone moments before the crash, as well as throughout the majority duration of the trip that night.

First, it can be identified that Rafaela Vasquez is an individual actor that represents one of many other individuals who also work as safety drivers for Uber ATG's data collection efforts. Therefore, the responsibilities of care for each of the drivers are identical. Each of the safety drivers are responsible "to monitor the driving environment and the operation of the automated driving system" as identified by the NTSB. Applying the four sub-elements of care, it can be shown that she was not attentive since she was distracted by her phone and could not become aware of the need, which was to override the self-driving and safely avoid the pedestrian. Since she was not attentive, she had no possibility to respond to the need, and therefore did not fulfill her responsibility. She had the competence of safely operating the vehicle, but it can be argued that since she was distracted, she did not have the competence to provide a good and successful care. Therefore, she was not competent. She was not responsive since she was an employee of Uber, and did not represent Uber in the best of light. In summary, she was not able to any of Tronto's four stages of care.

Another major actor that was involved in the accident was Uber ATG. They were entrusted by multiple actors to operate their self-driving vehicles on public roads safely. The permits were granted by the Arizona's state government, which entrusted the company to operate their vehicles on the public roads with safety. By proxy, pedestrians had an expectations of safety guaranteed by the state government. Therefore by the state entrusting Uber to operate safely on the road, Uber had a relationship with pedestrians to care for their safety. Applying the four sub-elements of the care, it can be shown that Uber ATG was attentive. To be attentive, the actor must be aware of the need. It can be shown that Uber ATG was aware of the needs of

pedestrians for their safety evident from the fact that Uber ATG operated the vehicles with a safety driver in place. It was also shown that the engineers and executives at Uber ATG were aware of the dire needs for the technical aspects of their ADS to improve in order to improve safety for all actors involved with public road usage. Specifically, it was known prior to crash that Uber ATG was aware of their ADS did not perform well when tracking bicycles, which was the edge case that ultimately led to the crash. While Uber ATG was attentive to the issues that were present, they were not completely responsible in responding and addressing the identified needs. Firstly, Uber ATG were well aware of the technical struggles of their ADS, but did not implement enough risk mitigating strategies to overcome the challenges. One example is the fact that only one safety driver was used per vehicle, instead of two (Templeton, 2019). While there are precedence in using only one safety driver per vehicle, such as for Waymo's self-driving project, many experts agree that Uber's ADS was not far enough along in technical development to risk using only one driver, since it was shown that Uber's ADS was struggling at a technical level. When Uber decided to move towards using one driver per vehicle, "some employees expressed safety concerns to managers" (Wakabayashi, 2018). Furthermore, it was revealed that Uber's training for drivers were not followed by every driver. For example, a driver was "fired after falling asleep at the wheel and being spotted by a colleague" (Wakabayashi, 2018). In addition, it was revealed that Uber removed the automatic collision brakes that came installed with the Volvo XC 90 (NTSB, 2019). These brakes would have provided a redundant safety system in case of unexpected behavior from ADS and the safety driver was not able to react appropriately in time, which was the exact scenario at the accident. For these reasons, and much more, it can be argued that Uber ATG did not act in a manner that would constitute a responsible actor. Uber ATG were also not acting in competent manner, as evident from their failure in

providing a good and successful care. Uber ATG was riddled with unsafe practices at many points within their operations including failure in hiring/training safety drivers, lack of success in overcoming the technical downsides that they identified, and general incompetence in safety protocols. In analysis of their responsiveness, it can be shown that they were not responsive since they did not consider the position of outside other actors. Firstly, they did not provide any mechanisms for the safety drivers to combat the fatigue of long hours with automation complacency, which ultimately led to the cause of the accident. In addition, they violated the terms of the permits that were granted for Uber in good faith by entrusting them to operate the vehicles in a safe manner on public roads. Lastly, the crash was partly caused by the fact that Uber's ADS did not account for the possibility of pedestrians jaywalking on streets with signs explicitly prohibiting crossing. By this, they have neglected to account for an expected human behavior, contributing to the cause of the crash.

While both the individual safety driver involved and Uber ATG did not fulfill the expected stages of care as identified by Tronto, it can be argued that Uber ATG bears a larger piece of ethical responsibility regarding this case. This is due to the fact that the safety driver was one individual actor out of many other individuals who all had equal responsibilities, but were all set up for failure at the hand of Uber's incompetence in addressing the safety concerns that they have identified from a both technical and systematic level. It is unfair that an individual who happened to fail in a system set up to fail is targeted to blame, when all of the reasons specific to the individual at the event of this crash was caused by Uber ATG.

Conclusion

The fatal crash of Uber's self-driving can be analyzed using care ethics framework, using Tronto's four sub-element of care to evaluate the level of care that was provided by each of the identified actors that were part of the network. Although in the aftermath of the crash, Uber ATG was able to legally settle out of court, as well as be ruled not criminally liable for the crash, through an analysis by care ethic framework as well as actor network theory, it can be shown that they did not provide adequate care at the four sub-elements of care, and is therefore morally responsible for the death of Elaine Herzberg.

Word Count: 3224

References

PRELIMINARY REPORT HIGHWAY HWY18MH010. (n.d.). Retrieved February 29, 2020, from https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

Wakabayashi, D. (2018, March 23). Uber's Self-Driving Cars Were Struggling Before Arizona Crash. Retrieved from https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html

Said, C. (2018, March 28). Uber puts the brakes on testing robot cars in California after Arizona fatality. Retrieved from https://www.sfchronicle.com/business/article/Uber-pulls-out-of-all-self-driving-car-testing-in-12785490.php

Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian. (2018, March 18). Retrieved February 29, 2020, from https://www.ntsb.gov/news/events/Documents/2019-HWY18MH010-BMG-abstract.pdf

ericavella2014. (2018, December 20). Uber's self-driving cars back on the road in Toronto after 9-month hiatus. Retrieved from https://globalnews.ca/news/4777626/ubers-self-driving-cars-toronto/

Lin, P. (2019, March 22). Who's at Fault in Uber's Fatal Collision? Retrieved February 29, 2020, from https://spectrum.ieee.org/cars-that-think/transportation/safety/reflecting-on-ubers-fatal-crash

Templeton, B. (2019, November 19). New NTSB Reports On Uber Fatality Reveal Major Errors By Uber. Retrieved from https://www.forbes.com/sites/bradtempleton/2019/11/06/new-ntsb-reports-on-uber-fatality-reveal-major-errors-by-uber/#d3a56221781d

Tronto, J. (n.d.). Partiality Based on Relational Responsibilities: Another Approach to Global Ethics. Retrieved from https://www.tandfonline.com/doi/full/10.1080/17496535.2012.704058

Sander-Staudt (n.d.). Care Ethics Retrieved from https://www.iep.utm.edu/care-eth/#SH1a

Callon. M. (n.d.) Actor Network Theory. Retrieved from https://www.sciencedirect.com/topics/neuroscience/actor-network-theory