An analysis of Uber's self-driving software:

The case of the first ever self-driving car accident

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

By

Jui Tao Tsai

April 2, 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.

Signed: Jui Tao Tsai

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

Introduction

In 2018, a Volvo car with self-driving technology integrated from Uber, in the state of Arizona, killed a pedestrian, Elaine Herzberg, who walked across a roadway. With further investigations conducted, they found that even though the detection system was running properly and it had no faults or diagnostic messages, it failed to further react to a jaywalking pedestrian as Uber had disabled the emergency braking system (Tech news, 2019). Consequently, the safety driver was not ready to respond to the pedestrian on time. It was the first-ever self-driving car accident that drew plenty of attention from the public. The vast majority of the reports blamed Uber for disabling Volvo car's emergency braking system (Uber disabled the emergency braking in the self-driving cars, 2018). That decision indeed led to tragedy; however, the view that technical problems are solely responsible overlooks several other aspects of the accident such as the unstable power dynamics between Uber and other actors as well as the importance of every single actor within the network. If we only focus on blaming the executive decisions of turning off the emergency braking systems, we may fail to consider the complete picture and the roles of each component.

Through the case of Uber's first self-driving car accident, I will apply the Actor-Network Theory (ANT) to argue that, rather than blaming various technical systems, several other factors led to this tragedy, such as the power structures between Volvo and Uber, the roles of test drivers and software engineers and their responsibilities. I will begin by outlining the structure and function of ANT and apply ANT to analyze different actors and the relationships among them. With the foundation in place, I will define the key human and non-human actors within the self-driving technology network by tracking the formation of its network. Then, with the context of the network, I will first demonstrate the unstable power dynamics between Uber and Volvo. Secondly, I will then point out the unfulfilled duty of one of the most critical actors, the test driver. Last, I will analyze the role of engineers to show why they were also responsible for this accident.

Background

The self-driving car technology was starting to develop rapidly due to the rise of various technologies such as image processing and machine learning. Also, different competitors such as Tesla and Lyft were striving to invent the first ever reliable self-driving cars (Autonomous Vehicles & Car Companies l CB Insights, 2020). The technologies were built with many hardware components such as radar, lidar, and camera image processing, and software components such as Computer Vision or Deep Learning. Uber was collaborating with Volvo cars while testing the self-driving cars (Uber unveils next-generation Volvo self-driving car, 2019). Additionally, there was always a safe driver inside the car to monitor the car in case the technology failed to react to unexpected events accordingly. In 2018, March 18, the first-ever self-driving car collision happened, and the pedestrian that was hit tragically died due to severe injuries. Such an accident prompted an investigation by the Tempe Police Department and National Transportation Safety Board (NTSB). The vast majority of the information regarding the details of how the software, driver, car, or pedestrian reacted in this accident can be found in this investigation.

Literature Review

A wealth of research exists analyzing autonomous vehicles in correspondence to the rapid development of self-driving technology. In this section, I will analyze two research papers that demonstrate two different perspectives on the fatalities caused by autonomous vehicles. The first paper discusses the case of the first-ever Uber autonomous vehicle car accident, and the second paper draws attention to the responsibilities for crashes of autonomous vehicles in general. Both papers emphasize mostly on particular actors such as the test driver, car manufacturers, or the technology companies. These work evade making judgments in the relationship between Uber, the technology company, and Volvo, the car manufacturer, as well as the duties and responsibilities of the test driver and the engineers.

In *The Wild, Wild West: A Case Study of Self-Driving Vehicle Testing in Arizon*a, Alexandra DeArman described that although the car's sensor did successfully detect Herzberg and her bicycle, the software determined it did not need to react right away, as it only recognized Herzberg as an object in the vehicle's path. Alexandra further suggests that the problem was what the broader system chose to do with that information. He notes that Uber intentionally designed and programmed the autonomous vehicle to have a higher false-positive allowance because of Uber's desire to develop a self-driving car that delivers a smooth and safe ride to its customers. Therefore, rather than react to the pedestrian sooner once the sensor recognized the pedestrian as an object, it waited until it was confirmed that the moving object was a person, and thus caused the accident to happen. While Alexandra addressed the effect of Uber's executive decision on the design of the autonomous vehicle, he is missing a bigger picture where Uber had too much power over Volvo, which allowed Uber to disable one of the most critical safety features in the car.

In *Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis*, Alexander Hevelke, and Julian Nida-Rumelin discussed what potentially causes autonomous vehicles to crash and who is responsible for it. In this paper, Alexander and Julian discuss the responsibilities of car manufacturers and users. First, car manufacturers such as Volvo are ultimately responsible for the final product despite that they are not in charge of developing the self-driving software. Additionally, it further notes that the drivers in autonomous vehicles should be responsible too since they took the risk of using the vehicle, knowing and accepting that it might cause accidents. It may be an excellent approach to force car manufacturers and drivers to be careful and responsible when they decide to be involved in self-driving technology. Their points of view, however, focused too much on individual actors instead of the power dynamics and relationships between different actors.

While there is undoubtedly much to be learned from Uber's first-ever self-driving car accident case, it also holds great value in thoroughly analyzing how different actors were responsible and what caused such an accident. The current body of research lacks the analysis of the power dynamics and relationships between different actors. This paper will not only provide complete analysis on the duties of different actors but also will utilize a framework to develop a rigorous process of how the network reaches stability to describe why the actor-network was unstable and led to the tragic death of Elaine Herzberg.

Conceptual Framework

The science, technology, and society (STS) concept of ANT proposes a practical framework for systematically characterizing roles within the self-driving technology network because it allows different heterogeneous components and their connections within the network to be isolated and analyzed. In order to analyze the actors, the power dynamics, and the relationships within the network, I will demonstrate relevant concepts from ANT within the scope of Uber's first self-driving car accident case. Throughout this paper, ANT will be consistent with the form proposed by Michel Callon. In Essence, ANT is a framework for deconstruction and critical analysis of the formation and function of various sociotechnical systems. ANT is a theory that simplifies heterogeneous elements with heterogeneous relationships into a cohesive network of human and non-human actors defined by their relative positions within the network (Callon 1987). In this theory, we are assuming that nothing outside of the system would affect the actors and their relationships. ANT also states that it is not the independent strength of actors that create power within an actor-network; it is the strength of the associations that relevant actors have with each other (Latour, 1986).

The process of simplification and association defines the conceptual actor-network, and it lays out the actual development of the actor-network that can be traced employing Callon's concept of translation (Callon 1986). Translation constitutes the different phases of a general process, where the identity of actors, the possibility of interaction, and margins of maneuver are negotiated and delimited (Callon, 1986). In essence, we can break down the concept of translation into four phases: problematization, interessment, enrolment, and mobilization. First, in problematization, a primary actor appears. problematization defines the problem at hand to be emphasized by the actor-network, and it identifies the necessary joint actors to recruit and sets itself as the "obligatory passage point" (OOP). Other actors must pass through OOP to establish a stable and mutually beneficial network. Secondly, in interessment, the network builder attempts to stabilize the identity of other actors by recruiting them into the network to perform certain roles. The recruited actors align their interests with the problem and OOP initially defined by the primary actor. Third, in enrolment, a set of roles that are defined and attributed to actors who accept them. One important note is that in order for the actor-network to be stable, all non-primary actors within the network should accept and carry out their assigned roles as intended. Last, in mobilization, the network builder takes up its roles as spokesperson for the network, and thus the network starts to function as a cohesive whole. If any actors within the network fail to perform the role assigned to it by the primary actor, the network will fracture and eventually fail (Callon, 1986).

I will use Actor-Network Theory to detect and track malfunctions within a heterogeneous actor-network. The concept of translation allows me to identify the actors within the self-driving actor-network, how the network builder, Uber, recruited other actors, and what caused the network to destabilize. I will first utilize the concept of the power structure in ANT to explain how Uber was too dominant as an actor that eventually destructs the sociotechnical network, then that the test driver was not carrying his duty, and lastly how the engineers may be an actor that was neglected but were responsible.

Analysis

It is necessary to reshape the public and self-driving companies' actor-network in order for the analysis to follow. First of all, we need to define the heterogeneous actors that are present within the network. I will lay out different actors derived from the National Transportation

Safety Board (NTSB) report that analyzed the entire self-driving car accident and pointed out each actor within the accident. These actors include (i) *Tech companies* (e.g. Uber, Lyft, Tesla, etc), the network builder that makes executive decisions on what self-driving technologies to develop and how they are tested, (ii) *developers* who directly implement the self-driving technology and design the testing for the tech companies, (iii) *self-driving software* that is developed by the engineers. (iii) *test drivers* who monitor the self-driving cars during testing (iv) *car companies* (e.g. Volvo, Audi, etc) that manufacture cars for tech companies to integrate self-driving software into, (v) *the public* that represents residents/people in general nearby the area where self-driving testing is conducted, in the case of Uber's first self-driving car accident, it was located in Arizona state, (vi) *state governments* that give permissions to tech companies to test their self-driving cars on road.

Before I start analyzing the entire actor-network, I will use translation to determine how the actor-network is formed and later on describe how the network in practice is unstable. The general form of self-driving actor-network can be described as below. The first moment of the translation is problematization. During this phase, the network builder, Uber, sees a market need for self-driving cars. Autonomous vehicles not only reduce the number of accidents that happen every year but also make people's lives easier. Based on the problem definition, Uber identifies that software developers will develop self-driving software, Volvo will offer the cars to integrate the self-driving software into, and the state government will allow such self-driving cars to be tested on the road, and Uber sets itself as the obligatory passage point (OOP).

In interessment, Uber recruits some other actors into the actor-network by aligning their interests toward the definition of the problem and OOP. To begin with, Uber has the vision to produce autonomous vehicles ultimately, and thus first recruits software engineers to develop self-driving software that enables completely hands-off and automatic navigation within cars. Then, Uber recruits Volvo, the manufacturer that offers cars for Uber to integrate the software into (Volvo Cars to supply Uber with up to 24,000 self-driving cars, 2017). Afterward, Uber recruits the state government to allow self-driving cars to be tested on the road in public. Finally, Uber recruits test-drivers to monitor self-driving vehicles and the public to interact with the vehicles.

In enrolment, hypothetically speaking, engineers, test drivers, the public, and the state government will embrace their allotted roles and establish a mutual association. In such a scenario, developers conduct research and analyze the most reliable technologies to implement and recruit safe and reliable self-driving software. Then, with the creation of self-driving technologies, the state government allows autonomous vehicles to be tested and drive on the road, and the public will interact with autonomous vehicles safely, assuming that other actors achieve their duties. Additionally, the test-driver will be in charge of monitoring the software to ensure that it is working correctly for an extra layer of safety. Under these conditions, the public will eventually utilize self-driving cars to go anywhere with autonomous vehicles and thus successfully avoid car accidents and can multitask while driving. Finally, Uber will solidify the connections within the network, and it allows seamless communications between engineers, the public, and the state government. Finally, it results in minimizing car accident rates and maximizing road safety through safe and reliable autonomous vehicles.

In order for me to analyze the power dynamics and control within the Uber self-driving network, I will present and analyze the power structures between different actors, what the duties of each actor were within the network, and why the lack of fulfillment in actors' duties led to destruction. First, power dynamics between different actors will show how the structure within the network is malformed. Secondly, the test driver failed to carry out her duty. Last, I will discuss why the engineers who directly implement the self-driving software were also responsible for the self-driving car accident.

Imbalance of Power Structure

While forming the self-driving actor-network, in the network builder, it was noted that in the phase of enrolment, the scenario that I described is optimal, meaning that it is the ideal outcome. However, in practice, there is an imbalanced structure during enrolment. Uber, which offered self-driving software, collaborated with Volvo, which provided the cars to integrate. Volvo cars were initially equipped with several advanced driver assistance functions. The systems included a collision avoidance function with automatic emergency braking, and functions for detecting driver alertness and road signed information (NTSB Report Suggests Uber's Backup Driver More at Fault Than Car in Fatal Crash, 2019). However, most of the functions above were disabled when the test vehicle is operated in computer control but is operational when the vehicle is operated in manual control. Thus, the self-driving system relies heavily on an attentive operator to intervene if the system fails to perform appropriately during testing. According to Uber, the emergency braking maneuvers were not enabled while the vehicle is under computer control to reduce the potential for erratic vehicle behavior. The vehicle operating is relied on to intervene and take control, and the system was not designed to alert the operator (NTSB Report Suggests Uber's Backup Driver More at Fault Than Car in Fatal Crash, 2019).

In this case, Volvo allowed Uber to decide whether specific system triggers are enabled or disabled. It showed that Uber had far more power over the actual vehicle than Volvo. As a result, the emergency braking system failed to work accordingly when needed even though the self-driving software successfully detected the pedestrian as a person 1.3 seconds before impact (NTSB Report Suggests Uber's Backup Driver More at Fault Than Car in Fatal Crash, 2019). From that, we can vividly see that the power dynamics between Uber and Volvo was not well balanced and thus the network destabilized.

Unfulfilled Duty of Test Driver

One of the reasons why the network destabilized is that the test driver in the autonomous vehicle did not fulfill her duty of focusing on the monitor and being aware of the situation. According to Uber, safety drivers play an important role in autonomous vehicle testing. They receive training to assume control when an onboard computer faces a situation where the self-driving car does not react to unexpected situations correctly. The conditions of driving change drastically on the road, and safety drivers must remain alert constantly (The state of self-driving car laws across the U.S., 2018). Uber has mandated that a safety driver always pays attention during the ride so that they can be in charge of the autonomous vehicle in unexpected situations (Uber self-driving crash 'mostly caused by human error', 2019).

In the case of Uber's self-driving car accident that happened on 18 March 2018, during an investigation conducted by the National Transportation Safety Board (NTSB), it revealed that the safety driver was not aware or paying attention to the software monitor. In the car was the safety driver Rafaela Vasquez, according to the investigators, had been streaming a TV show on her phone while "monitoring" the self-driving car. In fact, footage showed that Ms. Vasquez spent approximately 36% of the journey that particular night looking at the device (Driver mostly to blame for fatal self-driving Uber crash, federal safety board finds, 2019). Uber's computer software successfully detected the jaywalking pedestrian as an object, and it showed on the monitor screen 5.6 seconds before the impact (Serious safety lapses led to Uber's fatal self-driving crash, 2019), and Mrs. Vasquez was supposed to pay attention to it though the computer software did not alert her immediately after it recognized an object. Mrs. Herzberg was then alerted and realized that a pedestrian was crossing less than one second before the accident happened, and thus failed to react accordingly. From this, it is evident that the test driver did not execute her own duty properly as the last layer of safety.

In the enrolment phase in the translation that I described earlier, it showed that in order for the network to be stable and for the self-driving software integrated to the whole system safely and reliably, the safety driver has to accomplish his or her duty. However, it is evident that the test driver executed her duties properly as the last layer of safety. When one of the key actors failed to fulfill his or her role, the network deteriorated and thus the tragedy happened.

Regarding the Software Developers

Within the context of actor-network deconstructed above, a seemingly obvious counterpoint that failed the network was the power dynamics between Uber and Volvo as well as the responsibilities of the test driver and the software developers. Although the academic paper, *Enabling Pedestrian Safety using Computer Vision Techniques: A Case Study of the 2018 Uber Inc. Self-driving Car Crash* showed that the self-driving software could not be optimized further, and thus blamed the fault on other actors such as Uber for disabling the emergency braking system instead of the software engineers who developed the self-driving software. I agreed that the software has its limit at the time, and it was Uber's fault for disabling critical safety features; however, it does not mean that the engineers did not overlook some actors that interacted with the software. Although the software cannot be optimized any further, the engineers were responsible for the communication interface between autonomous vehicles and test drivers. In the footage of the car accident, it showed that Uber's computers detected Ms. Herzberg 5.6 seconds as an object before the impact (Serious safety lapses led to Uber's fatal self-driving crash, 2019); however, despite that such information was presented on the screen, there were no alerts until 1.3 seconds before the impact when the software identified the pedestrian as a person (Uber self-driving crash mostly caused by human error, 2019). Although the autonomous vehicle was supposed to stop even with the reaction time of 1.3 seconds, the human safety driver could not react to the pedestrian on time due to how late the alert was. In this case, while the engineers were designing the software system, they did not consider that human safety drivers were not able to react that quickly, and they implemented in a way that relied too much on purely the integrated system. Evidently, the engineers failed to consider all the actors involved and displayed an indifference to the quality of the system they were building. Additionally, it showed a lack of clear communication and connections between different actors, like software engineers and test drivers, which resulted in the failure of the network.

Conclusion

I have argued that in the case of Uber's first-ever self-driving car accident, it was not merely the faults of individual actors, but rather the power structure within the actor-network and the unfulfilled duties assigned to each actor. Thus, I used the Actor-Network Theory to demonstrate that we can analyze the connections and each role within the network. Given how the network is formed and how to reach the stability for the ideal outcome, I drew on each missing element that eventually led to the accident. I discussed how Uber, the network builder, had too much power over Volvo, the test driver failed to achieve her duty, and the engineers lacked the careful considerations for test drivers while implementing the software interface.

With my analysis, it allows people in the technical field to know that only focusing on the technical components may fail to analyze every dot and connection within the actor-network. With a solid framework that gives insight on all aspects of the network, we can understand better why accidents happened and what the mistakes that led to the accidents. If all the actors within the actor-network carry out their duties and maintain a power structure that is balanced and healthy, the self-driving technologies will keep growing and maturing until we reach the point where it is finally safe and reliable to drive autonomous vehicles.

WordCount: 3600

References

Autonomous Vehicles & Car Companies l CB Insights. (2020, February 24). Retrieved from https://www.cbinsights.com/research/autonomous-driverless-vehicles-corporations-list/

Callon, M. (1986). Some elements of a sociology of translation: The domestication of the scallops and the fishermen of St.Brieuc Bay. *In J. Law (Ed.), Power, action & belief: A new sociology of knowledge* (pp. 196-223). London: Routledge & Kegan Paul.

Callon, M. (1987). Society in the making: The study of technology as a tool for sociological analysis. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 83–103). Cambridge, MA: MIT Press.

DeArman, A. (2019). The wild, wild west: A case study of self-driving vehicle testing in Arizona. *Arizona Law Review*, 61(4), 983-1012.

Hawkins, A. J. (2019, November 6). Serious safety lapses led to Uber's fatal self-driving crash. Retrieved from https://www.theverge.com/2019/11/6/20951385/uber-self-driving-crash-death-reason-ntsb-dcouments

Hevelke, A., & Nida, J. (2015). Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science & Engineering Ethics*, 21(3), 619-630.

Karsten, J., & West, D. (2018, May 18). The state of self-driving car laws across the U.S. Retrieved from https://www.brookings.edu/blog/techtank/2018/05/01/the-state-of-self-driving-car-laws-across-the-u-s/

Latour, B. (1986). The powers of association. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge* (pp. 264–280). London: Routledge & Kegan Paul.

Lee, D. (2019, November 20). Uber self-driving crash mostly caused by human error. Retrieved from https://www.bbc.com/news/technology-50484172

McCausland, P. (2019, November 11). Tech news: self-driving Uber car that hit and killed a woman did not recognize that pedestrians jaywalk. Retrieved from https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

Pollard, N. (2017, November 20). Volvo Cars to supply Uber with up to 24,000 self-driving cars. Retrieved from https://www.reuters.com/article/us-volvocars-uber/volvo-cars-to-supply-uber-with-up-to-24000-self-driving-cars-idUSKBN1DK1NH

Randazzo, R. (2019, November 20). Driver mostly to blame for fatal self-driving Uber crash, federal safety board finds. Retrieved from https://www.azcentral.com/story/money/business/tech/2019/11/19/driver-fatal-arizona-uber-crash-mostly-blame-ntsb-report-finds/4232936002/

Shepardson, D. (2018, May 25). Uber disabled emergency braking in self-driving car: U.S. agency. Retrieved from https://www.reuters.com/article/us-uber-crash/uber-disabled-emergency-braking-in-self-driving-car-u-s-agency-idUSKCN1IP26K

Shepardson, D., & Reuters. Uber unveils next-generation Volvo self-driving car (2019, June 12). Retrieved March 1, 2020, from https://www.autonews.com/mobility-report/uber-unveils-next-generation-volvo-self-driving-car

Stern, R. (2019, November 21). NTSB report suggests Uber's backup driver more at fault than car in fatal crash. Retrieved from https://www.phoenixnewtimes.com/news/ntsb-report-uber-backup-driver-fatal-self-driving-crash-tempe-10453507