

Using Actor-Network Theory to Assess the 2018 Fatality of Elaine Herzberg

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

By
Hill Duggan

April 24, 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: **Hill Duggan**

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

Introduction

In March of 2018, 49-year-old Elaine Herzberg died after being struck by an Uber self-driving car that failed to perform an emergency stop; a US federal investigation later determined that the car's emergency stop system was disabled entirely. This incident was highly publicized, much to the detriment of the public perception of Autonomous Driving System (ADS) technology. When situations like these occur, society is faced with an ethical conundrum. Questions arise regarding who is to blame, what the next steps are, and how this affects the technology in question. Herzberg's death provides a picture-perfect example of the Problem of Many Hands (PMH), which is defined as "the occurrence of the situation in which the collective can reasonably be held morally responsible for an outcome, whereas none of the individuals can reasonably be held morally responsible for that outcome" (Jeske, 2015). The crash called for a slew of private and public investigations, the results of which I discuss in my analysis.

Suffice to say, the findings were damning for Uber; not only did the company face legal fire, but its stock also plummeted. However, it is also important to realize that many people's receptiveness to ADS technology *as a whole* were also greatly diminished. This simple statement may seem intuitively true, but the current literature fails to make this connection and see its gravity as it pertains to the larger Uber-ADS network. There is a wealth of ethical literature that gives investigative parties the tools to systematically (almost algorithmically) determine which entity must pay reparations---sadly, this is because the case of Herzberg's death is merely a drop in a sea of corporate ineptitude, albeit a highly publicized one. There is also more niche literature which explores people's potent emotional (affective) response to self-driving car accidents, e.g. Herzberg's death. These bodies of research seem inextricably linked in some way, but that link is largely lacking in the ongoing discussion. In this paper, I argue that current scholarly discourse

overlooks the critical role of public perception in the fate of a burgeoning technological network; I specifically look at the case of the Uber-ADS network, whose ostensible failure I aim to show was spurred by a loss of public trust. If we fail to understand the power of public perception in situations like these, we stand to face a reality where society's trust of ADS is permanently broken. Furthermore, we should seek to ensure that no similar network fails in the same, supremely costly way.

My continued reference to this complex system as a “network” is no accident or stylistic quirk. In this analysis, I employ Actor Network Theory (ANT) to examine the role of the public in the ostensible failure of the Uber-ADS network. The harrowing story of Herzberg's death provides lurid insight into the fluctuating power dynamic between the self-driving car researchers (Uber) and public consumers. In cases like these, where the constantly moving parts seem too complex to grasp, ANT can serve as a very powerful lens through which we may gain some understanding. Before I begin in earnest, I want to clarify that this paper does not claim to provide answers to these daunting ethical questions---rather, I simply aim to elucidate the questions we must ask ourselves in the wake of such a tragedy in simple hope that history will not repeat itself for the same reasons.

Literature Review

In the introduction above, I mentioned two visible classes of scholarly literature I found relevant to the Uber crash. In this section, I provide a brief overview of one specific peer-reviewed instance within each of these adjacent islands of research, as well as a brief synthetic interpretation of their significance with regard to my paper's claim. The first of these sources deals with the general concept of PMH, which can be seen in the Uber disaster. Again, PMH refers to the “Problem of Many Hands,” which I will explain further below. The second deals

with the nature of the Affect Heuristic as it pertains specifically to self-driving car accidents. Like PMH, the Affect Heuristic is a concept I also promise to elucidate.

The first source I employ is a scholarly review by Diane Jeske of a book titled *Moral Responsibility and the Problem of Many Hands* (heretofore simply dubbed *Moral Responsibility*). This book was written by Ibo van de Poel, Lambèr Royakkers, and Sjoerd D. Zwart; it was published in a 2015 Notre Dame Philosophical Reviews electronic journal. According to Jeske's review, *Moral Responsibility* seeks to address the problem of how our society *should* assign moral responsibility in cases where large groups of people cause external harm. This is the very essence of the preexisting PMH---the authors simply seek to synthesize a novel and robust ethical framework (primarily using such tried-and-true frameworks as deontology, consequentialism, and virtue ethics) which is tooled specifically to PMH.

The second source I have selected is an article titled "Machine versus humans: People's biased responses to traffic accidents involving self-driving vehicles." This article was published in a 2019 ScienceDirect journal, and authored by Peng Liu, Yong Du, and Zhigang Xu. This article narrowly focuses on the human predisposition to react more negatively to a crash involving a Self-Driving Vehicle versus one only involving a Human-Driven Vehicle. The article specifies and analyzes a couple of potential underlying reasons for this, which I will briefly denote. The first reason listed is "Humans vs. Machines," followed by an analysis of the rapidly changing relationship between human and nonhuman agents in today's technological landscape. The second reason is referred to as the "Affect Heuristic," another cognitive bias. In short, people tend to gauge the objective severity of a situation by the magnitude of their *affective response* (i.e. their emotional response).

Though neither of these articles explicitly deal with Elaine Herzberg's death, their content can be easily linked to it. The goal of this paper is not only to tie in these specific concepts to the crash, but to illustrate the link that is missing between them in modern scholarly discourse. To reiterate, my claim is that the damage to Uber's public reputation following the crash was a significant impetus in the network's failure. To cogently argue this claim is the ultimate goal of my analysis. However, before I begin constructing my main argument, it is critical that I further explain the conceptual framework I will use to do so.

Conceptual Framework

The ethical quandaries we face in the real world do not seem easily defined by the same Boolean logic engineers crave. On the contrary, they often seem to contain many parts whose motion is governed by some calculus beyond our understanding. To make matters worse, each "part" often seems to consist of countless sub-parts, which are often equally unpredictable. These parts can be human or non-human, so it seems too simplistic to assess them in the same way.

Fortunately, in the 80s, French scholar Michel Callon saw the necessity for a conceptual framework to assess the world's inherently (perhaps infinitely) complex processes without outright denial or lossy reduction of their aforementioned complexity. So, he developed Actor Network Theory (ANT) with his colleagues Bruno Latour and John Law (Muniesa, 2015). As I mentioned in the tail end of my introduction, ANT will serve as a vetted "lens" through which I view the moving parts of the 2018 Uber disaster.

As a framework, ANT "attempts to 'open the black box' of science and technology by tracing the complex relationships that exist between governments, technologies, knowledge, texts, money, and people (Cressman, 2009)." The name "Actor-Network Theory" explicitly specifies a very important dichotomy to understand: Actors vs. Networks. Actors are simply

human or nonhuman entities which act on each other in a certain way. These inter-actor relationships are specified by a constructed network. This is a short and accurate definition, but there are some important (and often confounding) aspects of ANT that are critical to understand.

First important note: many (if not most) useful actor-networks are “heterogeneous,” meaning they are composed of both human and non-human actors. This is not a hard rule, but is important to understand for this paper. Second: singular actors can *themselves* be interpreted as networks. A perhaps useful metaphor: an individual human being can be viewed as either “Johnny Depp” (a discrete actor), or a network of sub-actors working together (Johnny Depp’s brain, Johnny Depp’s heart, Johnny Depp’s legs). One can see a rabbit hole begin to form; we do eventually get down to quarks and bosons here. This is why the use of ANT is not an exact science, but rather a subtle art. One using ANT should seek to strike a perfect balance between *simplicity* and *usefulness* in their network representation; to get bogged down by the minutiae is to lose your reader, but to over-abstract your network is to often miss important details. Well-interpreted networks can still have many actors (sometimes dozens), with complex webs of connections. However, there are also very helpful network representations that only contain a small handful of actors; truthfully, only two connected actors are required for a true actor network.

A final concept to understand is that of a network-builder, which ever Actor-Network must have. The builder is simply the actor which drives the network---it is the impetus. One could easily say that the network-builder for the Johnny Depp actor-network I briefly mentioned earlier is Johnny Depp’s brain; the brain is what drives every other part of the body. In my analysis, I will map each of these terms to an Actor-Network of my own construction. One can view my claim as asserting that our current view of the Uber-ADS Actor-Network is incomplete.

To that end, I have not found any other scholarly research which views this specific case through the lens of ANT; this is why viewing a singular problem through multiple conceptual frameworks can be very helpful.

Analysis

Actor Specification

Before I begin representing and interpreting firsthand textual evidence from my scholarly articles, I must first lay the groundwork of my analysis by meticulously mapping out the Actor-Network; this is important, as it will be the very cornerstone of my argument. First, I will simply lend my network a convenient name (which I have already used in this paper): the Uber-ADS Network. Next, I will specify the actors within this network. One important note before I begin: there is no objectively correct way to construct an actor-network, only ways that lend themselves more or less to one's argument. Although this is a network representation purely of my own construction, I will attempt to justify each choice I make along the way.

Within the scope of the case I am analyzing, Uber appears to be a fine choice for a builder. As I pointed out in the Conceptual Framework section, one can easily slip down a rabbit hole here---it certainly would not be unreasonable to say Travis Kalanick, Uber's founder and former CEO, is the "true" network builder (Blystone, 2019). By the same logic, one could accurately say it is Travis Kalanick's *brain* which is the true impetus, or perhaps even the DNA which specified the construction of his brain. Uber is certainly a very large and complex entity, a roiling actor-network on its own. For the sake of simplicity and clarity of my argument, I will simply encapsulate it as its own single actor, which certainly contains the true network builder. In the end, my representation of the Uber-ADS will seem very simple, containing only 3 (very large) actors. I will explain why later.

The next actor I specify is not an amalgamation of humans like Uber; in fact, there is nothing human about it: it is an actor containing all self-driving cars, which I will dub the Vehicles actor. One could argue that only a very specific one of these Vehicles seems to be relevant (namely the one which struck Elaine Herzberg). One may also argue that only the Vehicles directly produced *by* Uber would be relevant. However, I posit that an actor which contains not one, but *all* of the self-driving vehicles (certainly *including* those in Uber's road-test program) will help paint a clearer picture of what caused the network to fail.

The final actor, and also the one I claim needs to be further explored, is the Public. This is an actor whose constituents are entirely human. Given the international attention this catastrophe received, I feel I can sweepingly say that all humans *not* contained within the Uber corporation belong to this actor. Of course, not all people in the world are aware of this incident, but I do not believe that precludes them from the umbrella term "Public." One could then take issue with my preclusion of Uber employees from the "Public." Again, this is a simple and arbitrary reduction for the sake of my argument---the argument should still work even if you take non-executive Uber employees as part of the Public actor, though I arbitrarily decided to keep them within Uber. As such, I specify that the set of humans within Uber and the set of humans within the Public are mutually exclusive and collectively exhaustive of the entire human population.

In sum, these are the actors of the Uber-ADS network: Uber, Vehicles, and Public. Many users of ANT find it useful to produce a pictographic representation of their Network(s), with lines denoting the relationships between them. Sadly, my network would make for a rather dull pictograph, as it would only be a triangle. So, for my next step, I shall *textually* explain the nature of the bidirectional connections between actors. After all, an actor-network is lifeless

without the denotation of these connections. Below I define and analyze the relationships between each pair of the 3 Actors: Uber vs. Vehicles, Uber vs. Public, and Public vs. Vehicles. Note that my use of ‘vs.’ simply indicates the existence of a relationship, though not *necessarily* one of opposition e.g. Ali vs. Frazier.

Uber vs. Vehicles

The connection between Uber and its Vehicles is the simplest, and also the one which I would argue garnered the most attention in the wake of the disaster. It is also the one least pertinent to my claim---I would be hard pressed to say anything novel about this connection in the Actor-Network. However, I will provide a brief overview of the relevant information. Uber *produces* Vehicles, specifically of the self-driving variety. Uber began its aggressive development of self-driving cars in 2015 when it poached 50 people from Carnegie Mellon’s robotics lab to populate its Advanced Technologies Group subsidiary (Lowensohn, 2015). Only a year later, Uber debuted its self-driving cars for select customers in Pittsburgh (Tascarella, 2016). This rapid and reckless expansion continued, with new consumer-available road-test programs opening in San Francisco and Arizona in the ensuing years. This all, of course, culminated in the death of Elaine Herzberg in Tempe, AZ on March 18, 2018. I will not spend much time discussing the or interpreting the 400 pages of documents that the NTSB released, which provided all of the lurid details of the crash. This is an area where current research is *not* lacking, so I shall skip unnecessary details. The following are a few of the compiled and summarized findings of the NTSB by an article from The Verge: Along with many other startling revelations, investigators determined that Uber lacked a formal safety plan. In devising its Autonomous Driving algorithm, it failed to account for commonplace driving scenarios like jaywalking. Furthermore, Elaine Herzberg’s death was merely the final and most severe instance

in a series of dozens of other crashes Uber's road tests had undergone (Hawkins, 2018). All I will mention of the aftermath is that Uber responded by publishing a 70-page safety report, making the case for the safety of self-driving cars over human-driven cars (Wakabayashi & Conger, 2018). During the ensuing legal proceedings, Uber reached an undisclosed settlement with the family of Elaine Herzberg, thus rendering the issue of liability a moot point (Neuman, 2018). In 2020, Uber regained its permit to legally test self-driving cars on public roads in California, with the presence of a backup driver required (Korosec, 2020). That is all I feel is relevant to say for my argument. Now that these two actors are connected, I can move onto the two relationships containing the public, which is where the bulk of my analysis will lie.

Uber vs. Public

This connection in the Uber-ADS network can be described generally as the public perception of Uber. By inspecting this relationship, I seek to showcase with evidence that Uber's reputation took a major hit following the 2018 fatality. Furthermore, I will ultimately explain why this is important for the failure of the Uber-ADS network.

A 2019 Reuters poll found the following: "Half of U.S. adults think automated vehicles are more dangerous than traditional vehicles operated by people, while nearly two-thirds said they would not buy a fully autonomous vehicle (Lienert, 2019)." While this poll data does not specifically regard public trust of Uber, the article summarizing and analyzing the data contains the following quote from Dan Sperling, UC Davis' director of the Institute of Transportation Studies: "At the moment, those responses are largely based on zero knowledge and zero experience, so it's mostly a visceral reaction to something they read about, like the (2018) Uber crash in Arizona" (Lienert, 2019). If people largely distrust the whole ADS industry, and people

recognize Uber as a part of the ADS industry, then by simple transitive property people distrust Uber.

To provide another example with a similar principle, I found recent poll results from the Edelman Trust Barometer, a well-reputed international trust and credibility survey. With regard to public trust of corporate CEOs: “63% [of participants] said [they distrusted] CEOs. The credibility of CEOs fell by 12 points this year, to 37% globally” (Harrington, 2017). Again, we can apply the similar transitive property argument to Uber if we assume Uber is one of the most globally recognized corporations. However, to minimize the use of assumption as a crutch for my argument, I found data from SensorTower Data Digest indicating that Uber was the 20th most downloaded app worldwide in the first quarter of 2019 (Briskman, 2019). Perhaps this lends credence to my position that Uber has substantial brand-recognition, but it could detract from my argument that the public has lost trust. So, I looked at the 2016 year-in-review data from the same website, and found that Uber was the 7th most downloaded app from the App Store worldwide 2.25 years prior (Nelson, 2017). It is an important reminder that correlation does not imply causation here, but it is interesting to note the contrast before and after Uber’s largest public scandals. I could delve into the plummeting quarterly reports, but I feel it is not necessary. The point I hope still stands, that Uber has very noticeably fallen in the public eye following the 2018 crash. One could easily make a very convincing argument (an argument I would likely agree with) that the 2018 crash is not the largest factor negatively impacting Uber’s reputation. It certainly was *not* the first or last of its faux pas; people have gone so far as to construct full timelines denoting Uber’s staggering number of public scandals (Carson, 2017). But still, to say it has fallen from grace *after* the crash is factually accurate. My argument will be

more complete after I explain the final relationship in the network, between the Public and Vehicles.

Public vs. Vehicles

These are the final unconnected actors in the Uber-ADS network. One can think of this connection as public receptiveness to ADS as a whole. With this section, I hope to round out my argument by providing substantial evidence that public trust of all self-driving vehicles was diminished by the Uber crash. Hopefully it is now becoming clear why I chose to include all self-driving cars in this actor---it seems safe to assume the public did not see Elaine Herzberg's death and think to themselves "I don't trust that *one* Uber self-driving car anymore! I will not ever ride in that specific car until the day I die." It is probably *less* safe to assume that the public did not think, "I don't trust any of *Uber's* self-driving cars anymore." In fact, I would argue this is a very reasonable camp to fall in. However, I want to try and mitigate the employment of assumptions in my argument, so below I provide an evidence-based argument that people statistically lost trust for *all* self-driving cars following the crash.

The Reuters poll I cited in the previous section is also relevant here: "Half of U.S. adults think automated vehicles are more dangerous than traditional vehicles operated by people, while nearly two-thirds said they would not buy a fully autonomous vehicle" (Lienert, 2019). Also recall the accompanying quote tying these results to a "visceral reaction to something... like the (2018) Uber crash in Arizona" (Lienert, 2019). This quote alone, even paired with the poll figures, could be dismissed as mere conjecture; again, correlation does not imply causation, especially when the poll is not specifically about Uber. For this reason, I will provide one more body of evidence and commentary.

The following quote is pulled from one of the scholarly journals I mentioned in my Literature Review section, the one regarding the Affect Heuristic: “People have a natural propensity to mindlessly apply social rules and expectations to nonhuman agents and therefore show the same social reactions in their interaction with nonhuman agents” (Liu et al., 2019). It is important to reiterate that the title of the article is “Machine versus humans: People’s biased responses to traffic accidents involving self-driving vehicles.” I will provide my own brief interpretation of the quote. The authors of the article are saying that if a human agent (let us say it was Johnny Depp) had come to Tempe, Arizona and murdered Elaine Herzberg for jaywalking, the public would have responded a certain way. Namely, we would have a strongly negative affective response, ideally putting Johnny Depp behind bars for a very long time. The use of a celebrity for this analogy is helpful because this would likely garner international attention in the same way the Uber car did. However, one would hope the public would not lose all faith in humanity just because Johnny Depp lost his marbles or has strong opinions on jaywalking. This is what the article is saying we effectively do with self-driving cars, however. This is the very nature of the Affect Heuristic.

Conclusion

I hope that in drawing the final edge of the Uber-ADS network, my claim finally seems reasonable. With this paper, I sought to provide an evidence-based argument calling for scholars to take a closer look at the role of the public perception in the failure of the Uber-ADS network. My intent was to be rigorous, though not at the cost of clarity. Furthermore, I hope I have successfully shown that the two “islands” of scholarly research detailed in my Literature Review (Problem of Many Hands & Affect Heuristic) are not too terribly distant, but rather close enough to be connected by a bridge, so to speak (per this metaphor, my argument aims to be the bridge).

My argument is fairly narrow in scope, as it only explicitly pertains to the failure of the Actor Network formed by Uber. However, I would humbly posit that my argument *is* significant, and one I hope people would ponder. It is perhaps especially relevant to those people who seek to drive the world forward into a driverless world. Uber is certainly not the only corporate entity developing and researching self-driving cars; Google, Tesla, General Motors, Toyota, and Nissan make up only a combined fraction. I could argue that Uber simply lies most centrally in the public eye for its near-comical volume of shortcomings and scandals. However, comical is never the right word when a life is lost to a scandal. Furthermore, I would assert that those in power at *any one* of these corporations need to think very carefully about how the public perceives them--to act so unethically like Uber did in this case has repercussions governed by the butterfly-effect, rippling far beyond its market share. Elaine Herzberg is ultimately 1 life lost to self-driving cars; we seem to ignore the 5-figure death toll from roadway fatalities that has plagued the US every year since 1918 (Holodny, 2016). Cars truly become safer every year, but that does not mean people become any less fallible. If these network-builders truly strive for a future with driverless cars (a future I could write another paper arguing is ultimately utopian), they must know that public trust cannot be so brazenly broken again. Otherwise, the whole ADS network may collapse permanently, and more lives will continue to be lost on the road.

Word Count: 4130

References

- Blystone, D. (2019, June 25). The Story of Uber. Retrieved April 23, 2020, from <https://www.investopedia.com/articles/personal-finance/111015/story-uber.asp>
- Briskman, J. (2019, May 15). Top Apps Worldwide for Q1 2019 by Downloads. Retrieved April 10, 2020, from <https://sensortower.com/blog/top-apps-worldwide-q1-2019-downloads>
- Carson, B. (2017, June 26). Uber's bad year: The stunning string of blows that upended the world's most valuable startup. Retrieved April 10, 2020, from <https://www.businessinsider.com/uber-scandal-crisis-complete-timeline-2017-6>
- Cressman, D. (n.d.). A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation.
- Harrington, M. (2017, April 11). For the First Time in 17 Years, People's Trust Declined in Every Kind of Institution We Asked About. Retrieved April 10, 2020, from <https://hbr.org/2017/01/survey-peoples-trust-has-declined-in-business-media-government-and-ngos>
- Hawkins, A. J. (2019, November 6). Serious safety lapses led to Uber's fatal self-driving crash, new documents suggest. Retrieved April 10, 2020, from <https://www.theverge.com/2019/11/6/20951385/uber-self-driving-crash-death-reason-ntsb-documents>
- Holodny, E. (2016, April 20). Traffic fatalities in the US have been mostly plummeting for decades. Retrieved April 10, 2020, from <https://www.businessinsider.com/traffic-fatalities-historical-trend-us-2016-4>

- Jeske, D. (2015, November 13). [Review of the book *Moral responsibility and the problem of many hands*, by I. Poel et al.]. *Notre Dame Philosophical Reviews*. Retrieved from <https://ndpr.nd.edu/news/moral-responsibility-and-the-problem-of-many-hands/>
- Korosec, K. (2020, March 10). Uber self-driving cars are back testing on San Francisco streets. Retrieved April 10, 2020, from <https://techcrunch.com/2020/03/10/uber-self-driving-cars-are-back-testing-on-san-francisco-streets/>
- Lienert, P. (2019, April 1). Americans still don't trust self-driving cars, Reuters/Ipsos poll finds. Retrieved April 10, 2020, from <https://www.reuters.com/article/us-autos-selfdriving-poll/americans-still-dont-trust-self-driving-cars-reuters-ipsos-poll-finds-idUSKCN1RD2QS>
- Liu, P., Du, Y., & Xu, Z. (2019). Machines versus humans: People's biased responses to traffic accidents involving self-driving vehicles. *Accident Analysis & Prevention*, 125, 232–240. doi: 10.1016/j.aap.2019.02.012
- Lowensohn, J. (2015, May 19). Uber gutted Carnegie Mellon's top robotics lab to build self-driving cars. Retrieved April 10, 2020, from <https://www.theverge.com/transportation/2015/5/19/8622831/uber-self-driving-cars-carnegie-mellon-poached>
- Muniesa, F. (2015, March 12). Actor-Network Theory. Retrieved April 10, 2020, from <https://www.sciencedirect.com/science/article/pii/B9780080970868850011?via=ihub>
- Nelson, R. (2017, January 13). Top Apps of 2016: Spotify, Line, and Netflix Led the Year's Biggest Earners. Retrieved April 10, 2020, from <https://sensortower.com/blog/top-apps-2016>

Tascarella, P. (2016, September 14). Uber debuts self-driving cars for Pittsburgh customers.

Retrieved April 10, 2020, from

<https://www.bizjournals.com/pittsburgh/news/2016/09/14/uber-debuts-self-driving-cars-for-pittsburgh.html>

Wakabayashi, D., & Conger, K. (2018, December 5). Uber's Self-Driving Cars Are Set to Return in a Downsized Test. Retrieved April 10, 2020, from

<https://www.nytimes.com/2018/12/05/technology/uber-self-driving-cars.html>