**ETHICS IN ENGINEERING: THE TROLLEY PROBLEM IN SELF-DRIVING CARS**

**Avery Peiffer (**[aep65@pitt.edu](mailto:aep65@pitt.edu)**)**

**THE SCENARIO**

Well, here I am. First staff meeting as a part of the engineering team. I’ve only been working here at rebU for a short time, but I’m really enjoying this self-driving car project, and I want to continue working on it. There is a very nice selection of donuts in this conference room. How many is it acceptable for me to take? I could try to stay in here after the meeting is over and swipe all of them. Okay, here’s our supervisor now; stop thinking about donuts. It’s time to get serious.

“Alright everyone, we’ve nailed our first iteration of self-driving cars. Their performance throughout a battery of tests proves that they are significantly safer than traditional driving. Only once has one of our cars been involved in an accident and that occurred when a human driver ran a red light and crashed into it in an intersection.

“We’ve received a lot of positive feedback on this project. Upper management wants us to continue developing these self-driving cars to the point where they can make decisions without a human present. I want you all to start thinking about how our system will approach those situations, and we’ll begin implementing additional software in the cars. Thank you all, and let me know if you have any questions.”

Wait, what? Hold on. What happens if one of our cars is going to hit a group of pedestrians and can’t stop itself? How does it respond to that? Would it endanger the lives of two passengers to save the lives of ten pedestrians? And you’re telling me that we have to make this decision before any car equipped with this software actually goes on the road?

**OUTLINE OF THE PROBLEM**

This issue is a modern variation on the Trolley Problem, a famous ethical dilemma that has provoked debate since its inception [1]. In the original Trolley Problem, a trolley is barreling towards five people tied to a track, unable to be stopped. You, an uninvolved bystander, have a choice: you can remain passive, allowing the trolley to continue and, likely, kill the five people, or you can pull a lever and divert the trolley onto a different track where only one person is tied down. At this problem’s center is the idea of value; the value of one life versus the value of five lives, and the value of one’s guilt either in pulling the lever or neglecting to act.

To define the hypothetical situation rigidly, I am forced to decide on how an autonomous car should respond in a situation in which neither choice is ideal, and where human override will do nothing to help. For example, the car may be driving on a one-way street when a group of pedestrians suddenly walk out in front of it. The software guiding the car calculates that it can either hit the pedestrians or quickly turn off the road, harming the car’s passengers. Because the car’s decision would affect the welfare of the entire public, the problem is thus defined in a manner that differentiates it from the usually self-contained constructions of this problem. Considering this problem allows me to explore and define the practical ethical principles to which I will have to refer in real-life engineering situations.

**FIRST STEPS: SAFETY AND COMMUNICATION**

When thinking about how best to address this problem, I first gathered information on how engineers have acted in similar situations. I referred to the ethics codes for the National Society of Professional Engineers (NSPE) and the Institute of Electrical and Electronics Engineers (IEEE). Most relevant in the general ethical standards set by the NSPE is its first canon, that engineers shall “hold paramount the safety, health, and welfare of the public” [2].

In most cases, this foundational tenet is in no way contentious because it is obvious when an engineer does not prioritize the public’s safety. We see this in the case study of the Kansas City Hyatt Regency Hotel, in which engineers Daniel Duncan and Jack Gillum failed to adequately maintain the hotel’s second and fourth floor walkways, leading to more than 100 deaths [3]. In my hypothetical, it is impossible to determine who exactly the ‘public’ is that I am sworn to protect. Am I obligated to ensure the safety of the passenger, who has paid for the car? Or should I prioritize the innocent pedestrians who would otherwise be the victims of an accident? Because there are defensible arguments for both perspectives, it is absurd to conclude a certain course of action based on this tenet alone.

To further clarify this issue, I turned to the code of ethics set by the IEEE, which was designed to govern situations more frequently encountered by electrical and computer engineers. I was able to develop more of a position based on the organization’s first standard: that electrical and computer engineers must “disclose promptly factors that might endanger the public or the environment” [4].

A clear and effective dialogue with the public is critical in my situation. It would be immoral to deceive the public about the risks of either choice by withholding information concerning the car’s decision-making process, regardless of what that process is. Additionally, since I would be obligated to act in the best interest of my employer whenever possible, if I withheld information from the public, the status of the entire company would be jeopardized were an accident to occur in one of its cars.

I found further examination of this principle in a case study of Intel’s Pentium microprocessor in 1994 [5]. The microprocessor, which had easily-detectable flaws that led to errors in basic computation, continued to be sold despite Intel’s awareness of the issue [5]. Only following a massive storm of negative publicity did Intel actually rectify the situation, offering to replace the flawed processor for all customers [5]. Creating a substandard microprocessor can be excused with human error so long as the company’s response to the issue is swift and equitable. Intel’s unethical actions were in its negligence in releasing to the public the details of the Pentium processor, and its refusal to fix the situation until it became apparent that their stock prices and profits would suffer from the incident.

However, it is important to understand that simply opening communication with the public regarding either action would not provide enough of a moral defense. Communication is important, but it cannot be the foundation of a proposed solution to an ethical dilemma of this nature.

**OTHER PERSPECTIVES**

It would be unrealistic to say that I, a freshman in college with zero real-world engineering experience, can understand all variables in this scenario and possess the knowledge necessary to make a decision. However, I will still try to use what I do know to draw conclusions. It is only logical that I would turn to others who I feel offer a valuable and novel perspective on the issue.

In this situation, I would first consult my mother. Knowing her, she would think critically about the issue in an abstract sense, without regard to the physical capabilities of the technology involved. We would likely engage in a discussion regarding the far-reaching consequences of each possibility. She might cite utilitarianism to argue for a case-by-case decision-making system where the respective populations of pedestrians and passengers would determine the car’s action. Or she might argue that the car should decide based on the impact the pedestrians and passengers have on society, giving an example where a car can protect either a highly-ranked public official riding alone or a group of pedestrians who do not wield any direct power in society. From discussion with my mother, I would gain a sense of understanding of the many perspectives from which the issue can be addressed, allowing me to make a decision with consideration of all variables.

While abstract conversation is helpful in thinking about hypotheticals, I would still need input on how to approach my ethical dilemma from a practical sense. For this I would turn to the husband of my first cousin, once-removed: Lane Albanese, director of engineering at a semiconductor company in Santa Clara, California. Lane has worked in computer engineering for over thirty years, and would be able to provide valuable perspectives on the practical implementation of the decision I make. He might temporarily suspend the moral dilemma to focus on how the nature of my company’s product would change based on the course of action I take. For example, if the car was programmed to prioritize pedestrians over passengers, the company would have to take extra measures to improve the car’s overall safety, so customers would feel confident they would never be placed in that situation in the first place. Lane’s experience with the various facets of working as an engineer would allow me to understand that my decision is not as black and white as it seems; for either action, there are compensatory actions that can be taken to minimize its consequences.

**THE MORAL DILEMMA: MY ETHICS, AS INFLUENCED BY PHILIPPA FOOT**

Though I have not yet completed any formal writing regarding the Trolley Problem, I intend to reason through this variation carefully in an attempt to arrive at a definitive decision. I think in its most rudimentary form, this problem deals with two factors: the group of people I am obligated to protect — either passengers or pedestrians — and the actual execution of that decision.

In conducting research for this assignment, I stumbled upon a line of reasoning for the original trolley problem that I found highly applicable to this variation. Philippa Foot, an Oxford moral philosopher, created the ideas of ‘negative duties’ and ‘positive duties’ to assess the original trolley problem [6]. She defines negative duties as “the obligation to refrain from harming others” and positive duties as “the obligation to actively do good” [6]. Furthermore, Foot reasoned that negative duties are “always more urgent and weigh more than our positive duties,” meaning one should prioritize not harming others over taking actions to actively do good [6].

In terms of Foot’s reasoning, the act of saving the lives of pedestrians is considered a positive duty, because it would require me to program an explicit action into the software of the self-driving car. Likewise, preventing the harm of one or more passengers is considered a negative duty, because ensuring the safety of passengers is an inherent characteristic of cars. These two duties cannot coexist given the nature of this problem. According to Foot’s reasoning, it would be my moral duty to protect the passengers of the car before I take any actions to save pedestrians’ lives.

In a vacuum, I agree with Foot’s line of reasoning and ultimate conclusion. However, it would be difficult for me, in a practical sense, to essentially sanction the endangerment of innocent pedestrians when there are no preferable alternatives. As a software engineer, I would be accurately predicting how I think the car’s computer systems should react in this situation. In saying that I think the car should protect its passengers by any means necessary, I am providing the software with intent that may result in the loss of lives. It can then be stated that I am technically committing involuntary manslaughter because I am aware my actions may lead to lives being lost [7].

**REAL-LIFE COMPARISON**

If this exact situation were to occur while a human was driving a car, I believe that either decision he or she makes is defensible. In the moment, the driver’s natural instincts take over as he or she makes a split-second decision. As long as it can be legitimately argued that the driver’s actions were ethical, the urgency of the situation outweighs any moral dilemma that could occur.

Parallels can be drawn between my hypothetical situation and the story of US Airways Flight 1549, the subject of the 2016 film *Sully*. Shortly after takeoff, the plane lost power to both engines when a flock of geese flew into its turbine, at which point pilot Chesley Sullenberger was forced to decide between crash-landing in the Hudson River, or attempting to glide the plane back to La Guardia airport [8]. Based on his prior experiences, Sullenberger felt that to attempt a return to La Guardia would be incredibly risky and could result in the deaths of everyone on board [8]. Sullenberger decided to land the plane in the Hudson, knowing that doing so would most likely ensure the safety of his passengers and crew [8].

Sullenberger’s decision follows the line of logic set forth by Philippa Foot. The urgent negative duty of preventing harm from coming to the people on board the plane far outweighed the positive duty of landing the plane so as to cause as little destruction as possible. Sullenberger’s decision was questioned by the results of several flight simulations, and some speculated that he should have attempted to turn back to La Guardia, but no one went so far as to argue that it was in any way immoral to land in the Hudson River [8].

**CONCLUSION: THIS IS NOT EASY**

The difficulty of this hypothetical is that I must predetermine the outcome of situations without regard for their specific context. The reasoning of Philippa Foot regarding this type of problem is very compelling, but would be difficult to actually implement. However, if I were somehow able to know if a self-driving car was facing these circumstances and I was forced to decide its course of action in that instant, I would not hesitate to act out of concern about ethical conundrums.

It is ultimately too difficult for me to make a definitive decision on what I would do in this scenario. Realistically, I would do everything in my power to make sure a self-driving car is never put in this situation, possibly by improving the car’s software to identify these circumstances before they develop. From this problem, it is reasonable to conclude that any software can never be expected to resolve ethical dilemmas. The computer system in a self-driving car should prioritize the protection of its passengers, but there must also be a human to aid in decision-making in serious and urgent scenarios. It would be interesting to further explore this problem from the perspective of Asimov’s “Three Laws of Robotics,” as it is not difficult to find direct contradictions that would occur [9].

My exploration of this topic has led to an interesting conclusion unrelated to the direct scope of the problem: for however ethically questionable a decision may be under normal circumstances, that sense of doubt is decreased by a significant magnitude under high-pressure situations, such as the one faced by Sullenberger on US Airways Flight 1549. Though it does not directly relate to the field of engineering, this is an interesting principle to apply to other facets of life. We often look at past situations with a high degree of hindsight bias, feeling as if there were definitively right and wrong courses of action. Given the controversial nature of the Trolley Problem and many real-life examples, the complexity and ambiguity of ethics are easily demonstrated.

As it applies to engineering, I now know from my exploration of this problem that I may be forced to make decisions in my career which can be retroactively seen as ethically questionable. But there is more to a decision than simply right and wrong. I must be sure to consider each situation I face in its own right before applying principles of ethics to lead to a decision. The codes of ethics provided by both the NSPE and the IEEE will help guide me to a general idea of how I need to resolve dilemmas as an engineer, but I cannot simply treat them as blanket commandments that will prove to be universally applicable. As an engineer, I must always remember to consider the context of a situation and, to a safe extent, trust my instincts.

**SOURCES**

[1] J. Dean. “The trolley dilemma and how it relates to ethical communication.” Accessed 11.18.2017. <http://www.trolleydilemma.com/>

[2] “Code of Ethics for Engineers.” National Society of Professional Engineers. 07.2007. Accessed 11.18.2017.

<https://www.nspe.org/sites/default/files/resources/pdfs/Ethics/CodeofEthics/Code-2007-July.pdf>

[3] “Hyatt Regency Walkway Collapse.” Engineering.com. 10.24.2006. Accessed 11.18.2017. <https://www.engineering.com/Library/ArticlesPage/tabid/85/articleType/ArticleView/ArticleID/175/PageID/199/Default.aspx>

[4] “IEEE Code of Ethics.” Institute of Electrical and Electronics Engineers. Accessed 11.18.2017. <https://www.ieee.org/about/corporate/governance/p7-8.html>

[5] C. Fleddermann. “Engineering Ethics Cases for Electrical and Computer Engineering Students.” Institute of Electrical and Electronics Engineers Transactions on Education. 08.2000. Accessed 11.17.2017. <http://ieeexplore.ieee.org/document/865202/>

[6] P. Foot. “The Problem of Abortion and Negative and Positive Duty: A Reply to James LeRoy Smith.” The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine. 01.09.1978. Accessed 11.19.2017. <https://doi.org/10.1093/jmp/3.3.253>

[7] “Offenses Under Title 18: 8.110 Manslaughter – Involuntary.” United States Courts for the Ninth Circuit Manual of Model Criminal Jury Instructions. 10.2013. Accessed 11.20.2017. <http://www3.ce9.uscourts.gov/jury-instructions/node/568>

[8] J. Paur. “Sullenberger made the right move, landing in the Hudson.” Wired online publication. 05.05.10. Accessed 11.20.2017. <https://www.wired.com/2010/05/ntsb-makes-recommendations-after-miracle-on-the-hudson-investigation/>

[9] “Isaac Asimov’s ‘Three Laws of Robotics.’ Auburn University. 2001. Accessed 11.20.2017. <https://www.auburn.edu/~vestmon/robotics.html>

**ACKNOWLEDGEMENTS**

This writing assignment was made possible by many of my peers in the Swanson School of Engineering, who have continued to motivate me to put forth my best effort in these writing assignments and in all areas of college life. I would like to formally acknowledge Eric Sherbocker, Daniel Stumpp, and Clay Countess for especially keeping me on track through a storm of midterms, projects, and writing assignments.

I would again like to thank Avocet Greenwell for her help in revising my writing assignments. Her feedback has helped me to consistently improve as a writer over the course of the semester, and I am very thankful for the time she devotes to my assignments.

Finally, I would like to thank the cast and crew of the Netflix series *Stranger Things* for providing me with a show that has consistently elevated my mood over the past few weeks. I find that I am most productive when I set short-term goals and rewards, and both *Stranger Things* and *Stranger Things 2* have been fantastic in that regard.

**ADDITIONAL SOURCES**

A. Narayanan, S. Vallor. “Computing Ethics: Why Software Engineering Courses Should Include Ethics Coverage.” Communications of the ACM. 03.2014. Accessed 11.15.2017. <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=6&sid=f31ce4b8-6eb8-4926-afb5-15ada9f1e20b%40sessionmgr120>

J. Stieb. “A Critique of Positive Responsibility in Computing.” Science & Engineering Ethics. 06.2008. Accessed 11.15.2017. <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=9&sid=f31ce4b8-6eb8-4926-afb5-15ada9f1e20b%40sessionmgr120>