# The Ethics of Self-Driving Cars

Joe Vogel – November 23, 2024

What is the first thing that comes to mind when you think of San Francisco? I wish it was still the Golden Gate Bridge, but somehow, for me, it's become self-driving cars. Specifically, the Waymos we see on nearly every street now. I can only describe their behavior as uncomfortable. Something about seeing these cars cruise around without a driver or passengers feels so alien. It makes me question whether the idea of self-driving cars is even ethical to begin with. But what about them is worrisome? This paper will explore several ethical concerns surrounding self-driving cars.

## 1 The technology

Over the past decade, rapid developments in technology have enabled self-driving cars (SDCs) to become increasingly mainstream. An overwhelming number of companies are taking advantage of the young market. Many of them are smaller, newer companies like Waymo, whose sole purpose is developing SDCs. However, most of the automaker giants, such as Mercedes, Tesla, and Toyota, are also tackling the technology with tens of billions of dollars [3]. Less than two months ago Tesla unveiled the *Robotaxi,* a driverless cab that you can own. While you aren't using it, it would offer its services to the public [2].

It's important to know not all self-driving cars are equal. There are six levels of self-driving classified by the Society of Automotive Engineers (SAE). The levels range from 0 to 5. The more relevant levels 2, 3, and 4 are described as "Partial", "Conditional", and "High" driving automation, respectively [4].

Level 2 cars are focused on Advanced Driver Assistance Systems (ADAS). They often support features like hands-free steering, parking assistance, lane centering, and assisted braking and acceleration [6]. An example of level 2 driving automation is Tesla's Autopilot. Level 3 cars can drive themselves without any driver interaction, but only in certain conditions (specific roads, clear daylight, certain speeds, etc.). Additionally, the driver must stay alert at all times, as the system can request the driver to take control at any moment. An example of level 3 self-driving is Mercedes-Benz's Drive Pilot [1]. A level 4 car can operate completely autonomously. They do not even require pedals or a steering wheel. However, level 4 cars still need to meet certain conditions, like specific roads. Waymos are an example of level 4 cars.

## 2 Ethical concerns

This paper will explore three main questions. First, "By which ethical framework should self-driving cars abide?" Second, "What should a self-driving car prioritize in emergency situations?" (e.g. child safety v. elderly safety, passenger v. pedestrian). And third, "Who should be responsible when self-driving cars cause accidents?"

It is apparent that there are many more ethical concerns surrounding these vehicles, but unfortunately they cannot all be covered in this paper. To list a few, however: Is the concept of self-driving cars ethical at all? Can we trust self-driving cars? Should we be worried about our privacy? If you were buying a self-driving car, would you be okay with your safety not being the car's number one priority? If you were a pedestrian, how would you feel if a self-driving car injured you to increase the chance of safety for its passengers? Do we value our autonomy over a lower death rate?

Would you be able to accept the death of a loved one who died because of a self-driving car, even though the nationwide casualty rate of traffic accidents was significantly lower? Only one thing is clear; none of these questions have clear answers.

## 3 Ethical frameworks

Let us first quickly go over the three major ethical frameworks: utilitarianism, deontology, and virtue ethics. Whether we know it or not, most of the decisions we make follow these frameworks.

“Will I get $5 worth of happiness from this coffee?” That’s utilitarianism. Originating from Jeremy Bentham, utilitarianism compares the pain and pleasure of a given situation. We quickly notice this concept is hard to apply to situations where certain elements do not have clear monetary value, such as a human life, or lying.

Deontology, originating from Immanuel Kant, follows the main principle of “act only on that maxim whereby you can at the same time will that it should become a universal law” [5]. Deontology states we shall not lie, we shall not cause harm, we shall not treat people as a means to an end. We shall act based on our actions’ intentions, not their consequences.

Lastly, we have Aristotle's virtue ethics. Virtue ethics is based on finding the “balance” in every action. To use confidence as an example, virtue ethics says not to be too rash or too cowardly, but be courageous. Nobody is born virtuous. We acquire intellectual virtues through experience, making everyone’s balance, or mean, different. Some actions have no mean and are always morally wrong, examples include theft and murder.

So by which ethical framework should self-driving cars abide? Currently, self-driving cars follow a form of utilitarianism, where each possible action has a weight. Staying in lane has low weight (optimal), swerving has high weight, hitting a pedestrian has extreme weight, the car is constantly choosing the option with the lowest weight [7]. But who is deciding these weights?

Professor of Philosophy Johannes Himmelreich at Syracuse University argues that passengers should be able to decide the ethics of their self-driving car via personal ethics settings (PES) [12]. With PES, passengers might choose to prioritize children over the elderly, or themselves over others. An argument against PES is the moral proxy argument, where we consider the engineers to be akin to doctors in the medical field. They know what is best for us, the car, and society as a whole. To counter the moral proxy argument, we can look at the results of MIT’s “Moral Machine” experiment [13]. They indicate that when it comes to sparing the youth over the elderly, France, Greece, and Canada are the top three advocates. Conversely, Taiwan, China, and South Korea believe strongly in saving the elderly [14]. The results show clear differences in ethical values across cultures. For that reason, I agree with Hemmelreich and believe a form of PES is the solution. Ultimately and unavoidably, PES is still some form of personalized utilitarianism.

## 4 Prioritization

Applying self-driving cars to the trolley problem gives us an entertaining thought experiment: What would a self-driving car do if it were faced with the situation of killing five people in its path, or changing course to kill just one person? A utilitarian car would change course to minimize “pain”. A Kantian car, however, would not change course, as it would be using the one person as a means to save five others. A virtue ethicist car would have a harder time deciding. What if the one person on the other track is the CEO of Waymo? Contextual factors could complicate the decision.

But according to Mahmood Hikmet, Head of Research and Development at Ohmio, the trolley problem is nothing more than a thought experiment. The likelihood of a self-driving car facing a true trolley problem is constantly reduced because of progressing technologies like constant component health monitoring, advanced sensor and vision systems, caution procedures in poor visibility conditions, and more [7]. Still, if a level 4 or 5 self-driving car was somehow faced with a true trolley problem, theory suggests the car would choose a utilitarian approach, a decision roughly 90% of us would also make [17]. So when we ask the question “what should a self-driving car prioritize?” I believe PES is still our best solution.

That is not to say that self-driving cars don’t need to react quickly in emergency situations, they absolutely do. It simply means the trolley problem is not as important as more prevalent problems, like the cars facing unexpected obstacles. In fact, we can look at a real-life example that occurred in February of 2024. A Waymo in San Francisco struck a bicyclist when they unexpectedly appeared from behind a garbage truck. The Waymo applied the brakes as soon as the bicyclist was spotted, causing only minor injuries upon collision [8]. Our focus should be shifted toward perfecting the car’s response to these situations before considering a trolley problem. Right now, the car will almost always choose to brake [7]. Swerving often introduces too many unpredictable factors, like losing control of the car. Self-driving cars do not like playing with chance, so they choose to brake. Perhaps with more time, the cars will have a better understanding of their capabilities.

## 5 Responsibility

Another relevant incident is the death of Elaine Herzberg. In 2018, while crossing a four lane road, Herzberg was struck and killed by an Uber test vehicle. It is difficult to fully blame the car in this case, as the driver was expected to take control in emergency situations. Video evidence later showed the driver on their phone seconds before the crash. As a result, Uber was not found criminally responsible, and the driver was given three years probation for negligent homicide [9].

In the US alone, over 40,000 people die every year from car crashes [10]. It is clear self-driving cars will never be able to guarantee zero vehicle-related deaths, but it is estimated that 80% of all collisions are due to driver distraction within three seconds of the accident [11]. It is undeniable that self-driving cars would eliminate almost all distraction-related accidents. They have more cameras than we have eyes, and they never even blink. So the real trolley problem is this —Do we allow vehicle death rates to remain higher while assigning responsibility to human error, or do we greatly reduce the number of deaths while accepting that robots will “kill” us?—

And the question still remains, who is responsible if a self-driving car causes an accident? Justices Sabine Gless and Thomas Weigend, along with writer Emily Silverman, have put much thought into this question. It doesn’t take long to realize it would be hard to criminally punish a robot. We can’t sentence it to prison as it has no free will, it follows its programming. Then do we punish the programmer? Well, it’s tricky. The programmer could be found liable if their development of the technology is deemed negligent. But when the programmer has followed all regulations, it is extremely difficult to assign blame to anyone [16]. In the case of Elaine Herzberg, we saw that Uber was nearly found liable, but in the end it was the backup driver who received the sentence. The ruling does suggest that if there was no backup driver, the company would be responsible. It is also worth noting Uber and the Herzberg family came to a private settlement outside of court [9].

In California, current laws can attribute fault to almost anyone. The automaker, software engineer, vehicle owner, driver, and passenger can all be found liable depending on the accident. Even the government themselves can be found liable in the case of insufficient regulations. For now though, there is a massive gray area surrounding the responsibility of self-driving cars in accidents. In my opinion, companies should be responsible for the products they produce.

## 6 Looking ahead

Jack Barkenbus, Senior Research Associate at Vanderbilt University, claims that the self-driving car industry is moving too fast for its own good [15]. Barkenbus provides survey data showing more than half of the US public is understandably worried about the development of SDCs. This emphasizes the lack of trust most people still have in these vehicles. In order to progress, it is crucial that we slowly show the world, through meticulous data-driven design and the utmost consideration of ethics, that self-driving cars will not only be safer than human drivers, but also have ethical values that reflect our cultures and beliefs.

Putting our trust in machines is not something new. I don’t see it as trusting a robot; I see it as trusting the engineers dedicating their lives to these projects. I trust that they are designing safe and reliable machines like they have done for so much of our history.

## References

1. Mercedes Benz. (2024). “DRIVE PILOT.” *mercedes-benz USA.* https://www.mbusa.com/en/owners/manuals/drive-pilot
2. Tesla. (2024). “We, Robot.” *Tesla.com.* https://www.tesla.com/we-robot
3. Baldwin, Roberto. (2020). “Self-Driving-Car Research Has Cost $16 Billion. What Do We Have to Show for It?” *Car and Driver.* https://www.caranddriver.com/news/a30857661/autonomous-car-self-driving-research-expensive/
4. Society of Automotive Engineers. (2024). “Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.” *SAE International.* https://www.sae.org/standards/content/j3016_202104/
5. Koehl, Patrice. (2024). “Deontology.” *UC Davis ECS 89L.* https://www.cs.ucdavis.edu/~koehl/Teaching/ECS089/Presentations/Kant.pdf
6. Society of Automotive Engineers. (2019). “SAE Standards News: J3016 automated-driving graphic update.” *SAE International.* https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic
7. Hikmet, Mahmood. (2021). “Self Driving Car Ethics - The Trolley Problem for Autonomous Vehicles.” *Youtube.* https://www.youtube.com/watch?v=30YiMc1k2Xw
8. Hawkins, Andrew. (2024). “Waymo driverless car strikes bicyclist in San Francisco, causes minor injuries.” *The Verge.* https://www.theverge.com/2024/2/7/24065063/waymo-driverless-car-strikes-bicyclist-san-francisco
9. Wikipedia. (2018). “Death of Elaine Herzberg.” *Wikipedia.* https://en.wikipedia.org/wiki/Death_of_Elaine_Herzberg
10. Moore, Timothy. Gollub, Heidi. (2024). “Fatal car crash statistics 2024.” *USA Today.* https://www.usatoday.com/money/blueprint/auto-insurance/fatal-car-crash-statistics/
11. California DMV. (2024). “Driver Distractions.” *State of California Department of Motor Vehicles.* https://www.dmv.ca.gov/portal/driver-education-and-safety/educational-materials/fast-facts
12. Himmelreich, Johannes. (2022). “No wheel but a dial.” *Springer Nature Link.* https://link.springer.com/article/10.1007/s10676-022-09668-5
13. MIT Media Lab. (2014). “Moral Machine.” *Massachusetts Institute of Technology.* https://www.moralmachine.net/
14. Hao, Karen. (2018). “Should a self-driving car kill the baby or the grandma? Depends on where you’re from.” *MIT Technology Review.* https://www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai
15. Barkenbus, Jack. (2018). “Self-driving Cars: How Soon Is Soon Enough?” *JSTOR.* https://www.jstor.org/stable/26597985?casa_token=_jgMcmBLxeIAAAAA%3At
16. Gless. Silverman. Weigend. “If Robots cause harm, Who is to blame? Self-driving Cars and Criminal Liability.” *University of California Press.* https://online.ucpress.edu/nclr/article/19/3/412/68679/If-Robots-cause-harm-Who-is-to-blame
17. Cloud, John. (2011). “Would You Kill One Person to Save Five? New Research on a Classic Debate.” *Time*. https://healthland.time.com/2011/12/05/would-you-kill-one-person-to-save-five-new-research