

**Programming Autonomous Vehicles to Balance Driver Safety and Public
Appeal with the Moral Responsibility to Minimize Fatalities**

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

By

Michael Laterza

May 1, 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: Michael Laterza

Approved: _____ Date _____
Rider Foley, Department of Engineering and Society

Introduction

As our society inches closer to welcoming in autonomous vehicles (AVs), there are many very important questions to be answered. There is clearly a wide range of potential benefits, from reducing accidents to increasing efficiency of vehicles, but AV developers have done little to alleviate many of the public's concerns. These concerns include some important details like pricing, design, and the actual ability of AVs to drive better than humans, but these problems should hopefully all be answered in the coming years. AV technology is going to continue to develop, and with numerous large tech companies funneling resources into their self-driving car projects, we can anticipate that high-level AVs will begin making their way into everyday traffic. And despite this seemingly inevitable conversion to AVs, there are a number of issues that can slow their transition as a part of mixed traffic, where AVs will have to share the road with human-driven vehicles.

What's more important to many eventual buyers is the car's programming, which, due to the nature of AVs, will be responsible for all decisions, taking the human element out of driving. Manufacturers and consumers seem to have completely different ideas about how this should work: people tend to want to have the option to drive, but companies are pushing for cars that have no driver control whatsoever, as that nullifies the point of removing human error from driving. And if ever a situation arises where a difficult decision must be made, the physical code must exist to decide it, whether the driver would agree or not. If this decision involves reacting to a life-or-death situation, people might rethink even buying an AV in the first place. As a group, autonomous vehicle manufacturers must find a way to balance several huge issues: saving lives, preserving morality, and maintaining enough public appeal to convince people to put their lives

in an unbiased algorithm's hands. This research will consider what the public desires in autonomous vehicles regarding features, safety, and the car's moral responsibility to protect lives.

STS Research Topic

Recent development setbacks and accidents haven't done anything to alleviate the public's general hesitation at the idea of self-driving cars. A recent crash in Tempe, AZ in 2017 led to a wave of angry public responses, ultimately forcing Uber to suspend their AV test program despite the fact that the AV involved was seemingly obeying traffic laws at the time (Bissell, 2018). Another survey from 2014 showed findings that agreed with this trend of distrust, stating that drivers were "very concerned" about riding in fully autonomous vehicles (Schoettle & Sivak, 2014 p. 12). Drivers were also found to be unlikely to buy an AV for a price any higher than non-AVs, and would expect to spend their time in an AV watching the road anyway if they'd even be willing to ride in the first place (Schoettle & Sivak, 2014 p. 17). Manufacturers aren't exactly giving the public any reason to change their opinions, either. Waymo and other manufacturers recently asked that the NHTSA quickly remove the restriction that cars must include a wheel and pedals to drive, despite not having self-driving cars that are currently safe enough for this change (Hawkins, 2019). This suggests that there is a massive disconnect between the goals of AV manufacturers and what the public wants.

On the issue of making the right ethical choice, it seems as though this dilemma is best examined through the lens of virtue ethics as opposed to utilitarian ethics. Virtue ethics is primarily concerned with doing what a person would feel is the most morally correct, while utilitarian ethics takes a purely material look at our actions. Here, the utilitarian solution is to just kill as few people as possible, but this is hardly an easy solution to sell to an already distrustful

public. Instead, a more innately human solution is needed, and AV manufacturers must agree to be held accountable for any decision that is made that does not have a bias for minimizing fatalities. In truth, no ethical framework can be used to fully encompass the goals of AV algorithms, and perhaps the best agreement we can come to is a deontological approach. This field of ethics consists of a set of ethical rules that cannot be violated (Goodall, 2014), and while the vast amount of possible ethical considerations in AV accidents is too big to create a ruleset for, it represents a more humane way of looking at it than pure utilitarianism.

As for the problems associated with public distrust, I point out some of the clear shortcomings of AV manufacturers. Their behavior and actions have hindered the social construction of technology (SCOT) for developing AVs. SCOT is a framework based on the idea that various relevant social groups can use the interpretive flexibility of technology to help it meet their desires, ultimately moving technological development forward, and this can be easily applied to the initial development of the automobile (Kline & Pinch, 1996). Unfortunately, AV manufacturers seem to want to remove any capability for humans to drive their vehicles, a move which not only makes development seem very one-sided but also potentially magnifies the problems AVs may face. The threats of security, hacking, and cooperation with other vehicles on the road – especially in a transitional period where only a portion of cars on the road are AVs – are even larger issues when the failsafe of human override is removed (Koopman & Wagner, 2017). As it currently stands, there is very little interpretive flexibility in AVs, as drivers and users are given little control of the vehicle, if any. It seems as though the only thing any relevant social groups can do is decide whether or not to buy them – clearly there needs to be more communication between manufacturers and their future consumers to ensure that AVs can sell.

Case Context

Perhaps the best way to properly explain the ethical dilemma that AVs are facing is to examine the case of the so-called “trolley problem” (Nyholm & Smids, 2016). In this hypothetical scenario, a person is tasked with either allowing 5 people to die by doing nothing or acting immediately to save those 5 people but ultimately killing 1 person who was not initially in any trouble. This exact scenario is very unlikely to actually occur – in fact, the trolley problem has its fair share of problems, as the outcomes aren’t indicative of any systematic sort of behavior. However, in a world where possibly billions of self-driving cars may soon share the road, some variation of this issue is bound to happen eventually (Bonnefon et. al, 2016). As a human, you may be forgiven for making such a rapid decision, but an AV’s manufacturer must make this decision far in advance and be ready to deal with the consequences of choosing life over life.

What is most significantly missing from the trolley problem is of course the risk associated with the passengers of an AV. The trolley problem poses no danger to the person making the decision, but if a vehicle must choose between killing its passenger or killing 10 people, this becomes an entirely different issue. According to a study of respondents from Amazon’s Mechanical Turk, 76% agree that AVs should sacrifice their passenger in this case, but they only scored their likelihood to buy such an AV at 19 out of 100, signifying a hesitation to actually adopt such a morally correct vehicle (Bonnefon et. al, 2016). It also appears as though drivers care far more about the “perceived usefulness” of AVs and trust in them than they do about the potential convenience AVs could bring (Choi & Ji, 2015 p.698). This idea of easier driving was a huge factor behind the rapid development of autonomous driving technologies,

largely driven by DARPA's initial push 2 decades ago (Bimbraw, 2015), but the public is reluctant to buy AVs for that reason alone.

Another important detail here is the concept of levels of vehicle autonomy, which causes some confusion for consumers who may have different ideas of what an AV really is. The levels range from 0 to 5, where 0 is no automation and level 5 indicates full automation (The 5 Levels of Autonomous Vehicles, 2018). Level 5 AVs have not nearly been developed yet, but this might be the only level deemed safe for everyday use in all situations. Levels 1 and 2 include cruise control, driver assistance, and some very basic self-controlled functionality – these are already in some vehicles currently on the road. It's levels 3 and 4 that are more likely to cause some confusion. A driver is still required behind the wheel for level 3 autonomous vehicles, as they can only operate driverless in ideal road conditions, and this is probably the level that is most likely to be on roads for the next 5-10 years. Level 4 autonomy is closer to ideal, as cars can be completely driverless, but even this could be fairly far off. It might help immensely if manufacturers can better relay this info to consumers who may not know what an AV might actually entail. For the majority of this project, however, I am referring to level 4 or 5 autonomous vehicles – the vehicles that can actually be considered truly “self-driving”. These won't be publicly available for some time still (especially not level 5), but the ethical questions that are brought up usually refer to these vehicles, since the car probably only has to make these decisions at or above level 4 automation.

Research Question and Methods

I analyzed these issues by addressing the following question: How can AV manufacturers and the general public agree on a policy for making AVs both ethical and desirable? There are

many other studies that have examined this issue using a variety of strategies, but surveys stood out as the most common way to gauge feelings on AVs. I developed a survey of my own to answer questions specific to what I want to discover. This questionnaire collects background information about the person taking it and how often they drive, then it asks opinion questions related to self-driving cars. One page of questions is about opinion on certain AV topics, such as confidence in driving one and perceived knowledge on the topic, and these allow for respondents to respond with one of seven answers ranging from “Strongly Disagree” to “Strongly Agree”. There are also a few extra questions to ask what people find most important about AVs and what makes them optimistic or pessimistic towards the future of self-driving cars. The final page of the survey asks questions about the ethical dilemma I propose:

A person is riding alone in a self-driving car. Suddenly, they come to a situation in which the vehicle is going to hit and kill a pedestrian who walked out onto the road. The vehicle could alternatively swerve into a wall and kill its passenger, sparing the pedestrian.

They are then asked to decide whether the car should swerve or hit the pedestrian or choose randomly, and the question is repeated again for 5 pedestrians. These are both then asked again, but instead respondents are asked what they believe the manufacturers will program the car to do. There are also 2 questions to close the survey asking if this unlikely scenario changes their view on AVs.

In addition, I sought more complete answers beyond just multiple-choice responses, so I conducted a few brief interviews to obtain more detailed results. Many of the questions are similar to the survey, but they are left more open-ended. In particular, the ethics questions are worded different to ask if a utilitarian policy should apply (i.e. killing the least people) and how the scenario changes if we add people to the road or the car. I also have a longer question asking

participants to list several things that make them like or dislike self-driving cars, and a question asking if they have faith in AV manufacturers designing cars that would act similarly to them.

A total of 57 participants took part in the survey, among whom 75% identified as male. 77% of those surveyed were between the ages of 18 and 24, which restricts the ability to draw conclusions about age's effect on AV perception. The data was collected through Google Forms and exported into Google Sheets, where I analyzed the data and constructed figures. I also conducted 6 of the short interviews on a variety of interviewees. This data was collected on a basic text document to examine and exposed some interesting reasoning behind some of these AV opinions.

Results

As a whole, my research suggests that potential AV purchasers don't have much faith in AV manufacturers when it comes to both desirability and ethics. Almost half of the respondents assumed that manufacturers would program AVs to behave differently than they themselves think it should, and the questions that pertain to design choices have answers that also contradict what is being currently developed. Young people are likely the future market for AVs, yet my survey of mainly people under 25 shows trends that should alarm manufacturers about what this generation wants in their vehicles. The interviews showed similar results, with most respondents showing more concern and distrust than optimism, and they also agreed on a few things that could help manufacturers earn the public's approval in the future.

In total, 42% of participants indicated that they agreed slightly that they are well informed about AV developments, with another 14% agreeing more strongly than that (Figure 1). Only 30% disagreed with that statement, which shows that most participants are paying some attention to this issue. Also, only 28% disagreed with the statement that they could see themselves owning a self-driving car, indicating slightly more optimism about using AVs than seen in previous studies (Figure 2). Half of the interviewees also agreed that they would use an AV in the future and feel safe doing so, although one of them indicated that they wouldn't buy an AV "until everyone else does". In addition, only 2 respondents picked "Disagree" or "Strongly Disagree" when asked if they would feel safe in a self-driving car (Figure 3). Far more answers were clustered around "Somewhat Agree", showing increased faith in the promised safety of AVs. These initial results fit a gradual trend toward more public understanding of AVs and the capabilities they may have in the coming years, and less blind distrust of innovative technology as this future nears.

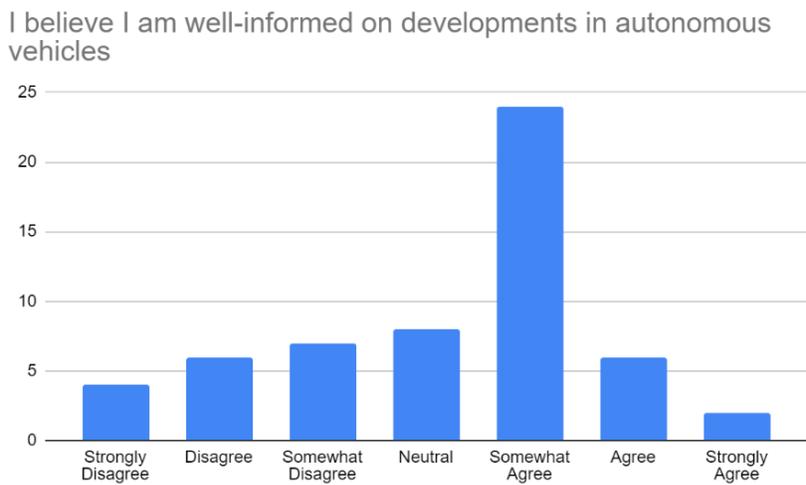


Figure 1. Agreement with statement about how informed participants are about AVs

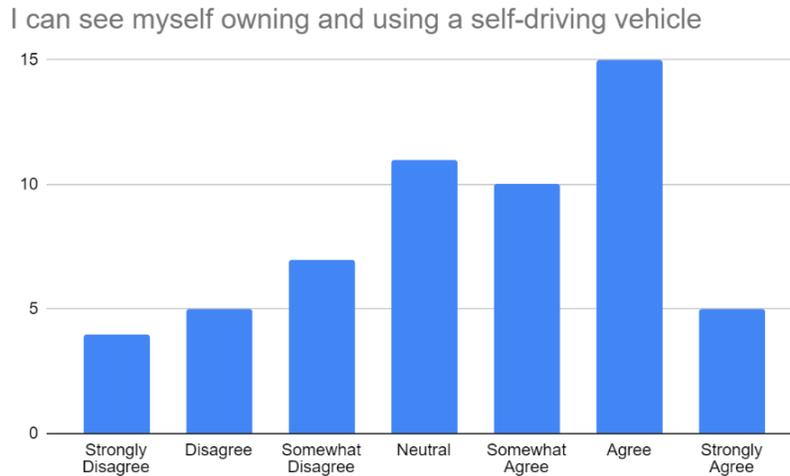


Figure 2. Agreement with statement about eventually owning a self-driving car

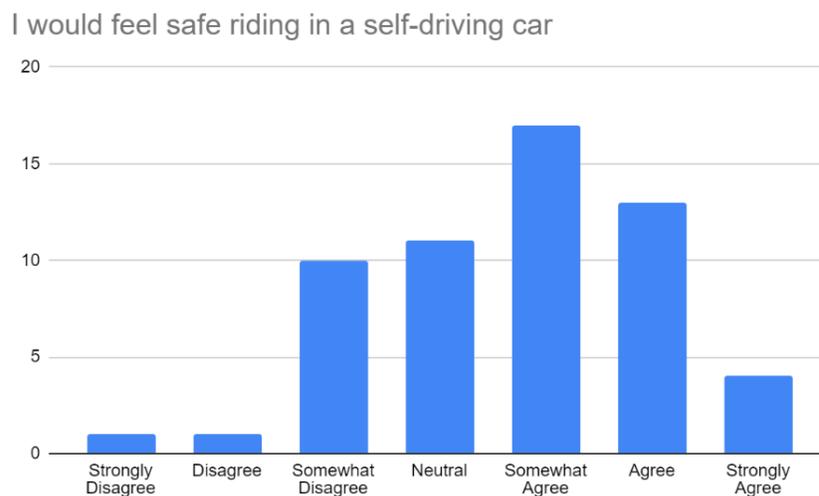


Figure 3. Agreement with statement about feeling safe in a self-driving car

These trends show a positive outlook toward AVs, but once the survey questions begin outlining some of the details of how self-driving cars will enter the country’s roads, the results begin to show less optimism. Mixed traffic consists of normal vehicles as well as self-driving vehicles, and the survey participants seem very divided about how safe they’d feel using an AV under this condition (Figure 4). Most of the results clustered around “Neutral”, with a lower average agreement than the participants indicated when traffic was not specified. In an even

more lopsided trend, 49% of participants either “Disagree” or “Strongly Disagree” that they would ride in an AV with no option for human control, and only 21% agree any amount (Figure 5). In another question, a massive 82% answered that they would prefer to have a driver takeover feature, mainly to have the freedom to drive when they choose. This would allow for far more interpretive flexibility in vehicles, but it seems to be at odds with the goal many manufacturers have for their AVs once level 4 or 5 autonomy is commonplace, as vehicle-to-vehicle communication would make a human driver detrimental (Hancock, 2019). In addition, the results skew even further when examining respondents who drive a car at least once a week, as 31 of those 39 respondents disagree with the statement and not a single one said they “Strongly agree”. All but one of the interviewees agreed that they wouldn’t want an AV without driver takeover, and even that one person indicated that they would only buy that if everyone’s AVs had to be that way. Three of the interviewees also indicated that a drawback of AVs would be losing the freedom to drive however they choose, and two believed that traffic would be slower since the cars might drive at a lower speed than they would, or even “below the written speed limit”.

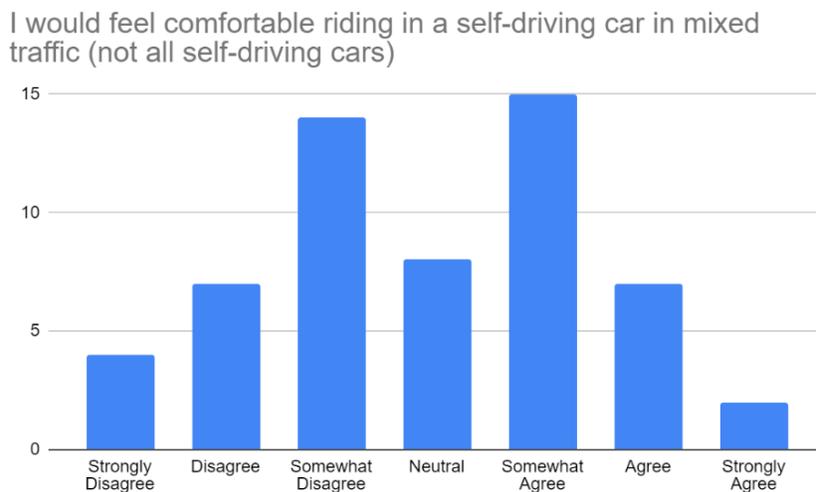


Figure 4. Agreement with statement about comfort riding in an AV in mixed traffic

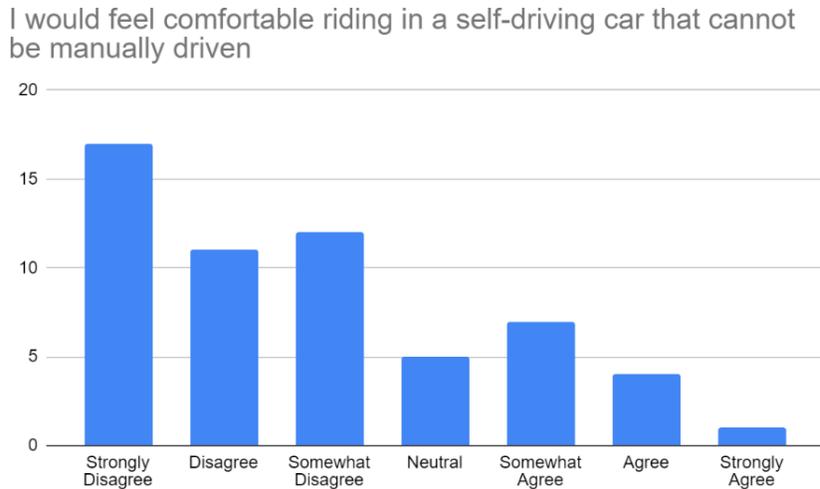


Figure 5. Agreement with the statement about comfort riding in an AV that a human cannot drive

This is all despite the survey’s participants agreeing that self-driving cars will probably reduce accidents once they reach their full potential (Figure 6). Besides the 17% that either disagreed slightly or felt neutral on that statement, everyone else (and 4 of the 6 survey respondents) showed faith that AVs can eventually reduce accidents, in line with a previous study of Pittsburgh residents (Penmesta et al, 2019). The few survey respondents and lone interviewee who did support removing driver takeover from AVs all understood that reducing accidents caused by other drivers is the reasoning behind such a decision. 58% of all car accidents involve only one vehicle (Beltz, 2020), so maybe that helps justify this contradiction between wanting to drive manually and thinking AVs can drive better than most people. People are confident in their own personal driving and want the ability to drive if they please, which is a sentiment that AV manufacturers might soon need to understand. When asked how they think they would spend their time in an AV, more than 50% chose the option about watching the road, which seems like it would be pointless if they cannot drive (Figure 7). Additionally, this jumps to over 60% of those who drive at least once a week, which might reflect that they are used to driving and watching the road and may not be able to envision doing something else. It has also

been proposed that vehicles could ease passenger fears by showing a visualization of what the car senses, which may lower this number (Hutson, 2017). Regardless, this trend of watching the road has been similarly high in other studies that asked about what people would do in an AV, and there is no way to tell yet if this will remain the case once more people ride in AVs: the only people currently riding in AVs are testers, who are supposed to at least partly have eyes on the road.

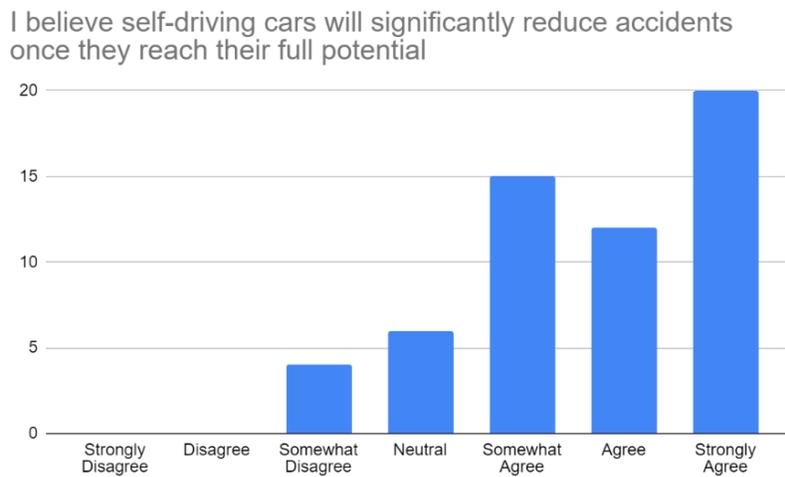


Figure 6. Agreement with statement about confidence that self-driving cars can outperform humans

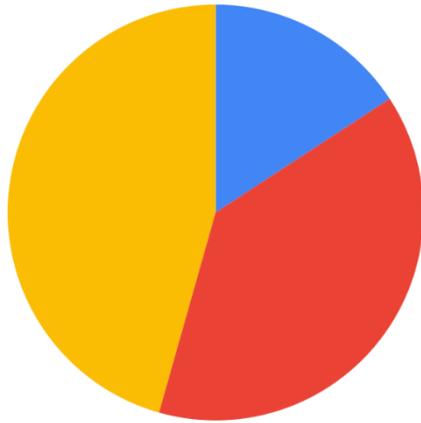
How participant would spend time in an AV	Count	Percentage
Watching the road	32	56.1%
Watching movies or TV shows	7	12.3%
Using phone/device	12	21.1%
Reading	3	5.3%
Other	3	5.3%

Figure 7. How participants believe they would spend most of their time in an AV

As for the questions related to ethics, there was not much of a consensus on how vehicles should handle a dilemma like the one provided in the survey. 46% think the car should choose who to kill at random, 38% think the car should hit the pedestrian, and only 16% would

recommend sacrificing the passenger (Figure 8). The interviewees were evenly split between the 3 choices, but it is significant that 4 of them asked whether or not the car or pedestrian were at fault for this scenario, indicating that they would opt to spare the person who is not at fault. When asked what they believe the car will be programmed to do, the answers shift slightly toward hitting the pedestrian (now 54%), possibly reflecting manufacturers' perceived attempt to reduce damage to their vehicles or make it easier to sell these vehicles. The interesting trend beneath this is that 24 of the 57 participants (42%) put down different answers for these 2 questions, showing that many respondents think the vehicles will act differently than they should. However, when the number of pedestrians is increased to 5, the results for both charts become nearly identical, with about 50% believing the AV will sacrifice the passenger to save more lives. This utilitarian decision probably feels like the natural choice for some, and all 6 interviewees agreed on this choice after picking all 3 answers when only 1 pedestrian was involved. And after this scenario and its repercussions were revealed, only 26% picked either "Yes" answer when asked if they would ride in an AV that could potentially choose to sacrifice them for the pedestrians (Figure 9), far lower than the percentage of agreeing answers in the initial question about riding in AVs from Figure 3. Among frequent drivers (once a week or more), this drops to 19%, likely because they currently feel as though they are safe enough drivers to either avoid this situation or solve it on their own.

In this scenario, which do you think the car SHOULD be programmed to do?



In this scenario, which do you think the car WILL be programmed to do?

- Swerve into the wall (kill its passenger)
- Hit the pedestrian (kill 1 other person)
- Choose randomly

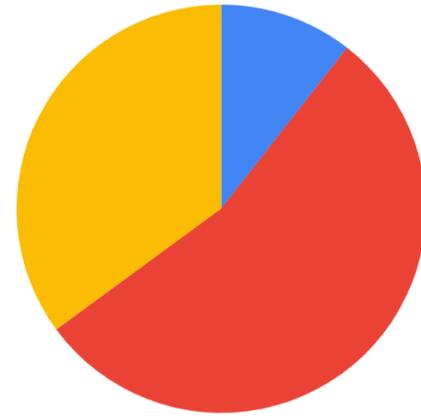


Figure 8. Pie charts for how people think an AV should – and will – react to the given scenario

If AVs sacrifice passenger in this scenario, do you still feel safe?	Count	Percent
Yes, this scenario will probably never happen	8	14.0%
Yes, I understand this design decision	7	12.3%
Maybe, I want to have driver control for chance to make decision myself	15	26.3%
No, and I already wouldn't have felt safe before this	10	17.5%
No, I would much rather make this decision myself	17	29.8%

Figure 9. Table of responses to question about riding in an AV that might sacrifice passenger if needed

Discussion

Many of the studies I referenced had results similar to mine, apart from those showing slightly less optimism toward driving or buying AVs. Looking at the years in which these previous studies were conducted, there does appear to be a subtle trend toward increased AV favorability, but this should be expected as advanced technologies like this become more commonplace in society. A 2015 study found a trend of significant concern and general lack of excitement towards AVs in studies on the topic from 2013 and 2014, but then found significantly more optimistic results just a year or two later when surveying a similar population (Kyriakidis

et al, 2015). My results fit this part of an upward trend toward accepting AVs, and also doesn't stray far from previous studies when it comes to distrusting manufacturers.

The ethical side of this problem isn't so simple either, as the results show. There are many confounding variables that make most people answer those ethics questions without fully considering their choice, and this was hard to reflect in a survey. For instance, a whole other argument can be asked about whether or not manufacturers should – or even can – be held accountable for deaths in AV accidents (Hevelke & Nida-Rümelin, 2015). This would obviously weigh heavily on the manufacturers' decision in this "forced-choice situation", which many believe AVs can totally avoid, or at least mitigate to the point where a person can feel confident that they will never experience it (Fleetwood, 2017). The survey anecdote did not go into too much detail on the vague underlying factors that could steer the ethics answers one way or another, in an attempt to keep the answers as unbiased as possible. This helped ensure that neither the vehicle or the pedestrian is doing anything wrong in the scenario, which could have heavily skewed the results toward whichever side is "in the right". The utilitarian option – killing one person to spare five - was more popular once one side of the scenario had more people involved, but the interviewees' answers indicate that the party with more virtue should survive if the number of deaths cannot be altered. There isn't a concrete answer choice to reflect deontological ethics, but programming with these ethics in mind would be able to reflect this balance between two different ethical choices.

There is room for expansion on this topic in future work. A larger and more diverse sample size would immediately help increase my confidence in my results and allow for more trends to appear based on the demographic information, and a longer survey could have eliminated the need for the additional interview. More intricate questions about certain AV

capabilities could prove useful, similar to a study showing which features people found likely to work (Bloom et al, 2017). I would recommend adding more detailed questions about the cost increase a person would be willing to pay for a self-driving car, as my questions on the topic found no conclusive result and a previous study did find a decline in favorability when given a price (J.D. Power, 2012). I'd also adjust the number of pedestrians or passengers in the ethical scenario and add a question to examine the perception of how soon this technology will be available. Also, I find that it may be revealing to repeat this survey for a few consecutive years on a similar audience, in order to find more concise evidence of a gradual shift in favor of buying AVs.

This research will help advance my development knowledge regardless of whether or not I wind up working on autonomous vehicles myself. In that scenario, I would know to be as transparent as possible and see myself as a consumer as well, but the overall trend holds true for any programmer or even any engineer. It helps to communicate well with clients and consumers since bold, innovative technologies aren't going to infiltrate our everyday lives without meeting the requirements set by the public. The principles of SCOT can apply to many technologies, so I will more than likely run into this framework again.

Conclusion

Even though this data was more recently obtained, the trends agree well with the conclusions drawn from previous studies. There is a general sense that manufacturers aren't allowing for any interpretive flexibility in autonomous vehicles, and there is no obvious choice for how to solve a life-or-death ethical dilemma. Many people see self-driving cars as an

inevitable future, but as with any new emerging technology, they need to be developed well in order to stick around. In the next 5-10 years, these cars will start hitting the roads in greater numbers, and more research will be needed to see if manufacturers have actually done anything to alleviate any of the public's concerns. This technology has the potential to save tons of lives and improve upon one of the most widespread form of transportation in the world, but if it is not designed or marketed in a way that appeals to the public, it may never become a reality. People need to stay well-informed about AVs, and it is vital for AV companies to continuously gauge public opinion in order to decide their next moves. Companies have put humungous investments into self-driving technology: now they just need to sell it to uncertain consumers.

References

- Beltz, B. (2020, January 10). 100 Car Accident Statistics for 2019. Retrieved March 10, 2020, from <https://safer-america.com/car-accident-statistics/>
- Bimbraw, K. (2015). Autonomous cars: Past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology. In *2015 12th International Conference on Informatics in Control, Automation and Robotics*, 1(01), pp. 191-198.
- Bissell, D. (2018). Automation interrupted: How autonomous vehicle accidents transform the material politics of automation. *Political Geography*, 65, 57-66.
- Bloom, C., Tan, J., Ramjohn, J., & Bauer, L. (2017). Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, 357-375.

- Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573-1576.
- Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, 31(10), 692-702.
- Fleetwood, J. (2017). Public health, ethics, and autonomous vehicles. *American journal of public health*, 107(4), 532-537.
- Goodall, N. J. (2014). Machine ethics and automated vehicles. In *Road vehicle automation* (pp. 93-102). Springer, Cham.
- Hancock, P. A. (2019). Some pitfalls in the promises of automated and autonomous vehicles. *Ergonomics*, 62(4), 479-495.
- Hawkins, A. J. (2019, August 30). Self-driving carmakers urge regulators to whiff the steering wheel out the window. Retrieved October 17, 2019, from <https://www.theverge.com/>.
- Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: an ethical analysis. *Science and engineering ethics*, 21(3), 619-630.
- Hutson, M. (2017). A matter of trust. *Science*, 358(6369), 1375-1377
- J.D. Power (2012, April 26). *Vehicle owners show willingness to spend on automotive infotainment features* [Press release]. Retrieved from <https://www.jdpower.com/sites/default/files/2012049-uset.pdf>
- Kline, R., & Pinch, T. (1996). Users as agents of technological change: The social construction of the automobile in the rural United States. *Technology and culture*, 37(4), 763-795.
- Koopman, P., & Wagner, M. (2017). Autonomous vehicle safety: An interdisciplinary challenge. *IEEE Intelligent Transportation Systems Magazine*, 9(1), 90-96.

- Kyriakidis, M., Happee, R., & de Winter, J. C. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation research part F: traffic psychology and behaviour*, 32, 127-140.
- Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem?. *Ethical theory and moral practice*, 19(5), 1275-1289.
- Penmetsa, P., Adanu, E. K., Wood, D., Wang, T., & Jones, S. L. (2019). Perceptions and expectations of autonomous vehicles—A snapshot of vulnerable road user opinion. *Technological Forecasting and Social Change*, 143, 9-13.
- Schoettle, B., & Sivak, M. (2014). *A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia*. University of Michigan, Ann Arbor, Transportation Research Institute.
- The 5 Levels of Autonomous Vehicles. (2018, March 5). Retrieved February 1, 2020, from <https://www.truecar.com/blog/5-levels-autonomous-vehicles/>