Matt Ritchie
CS 456
Final Report

Introduction:

A predominant ethical issue with the development of self-driving cars is what to program the vehicles to do in the worst-case scenarios. If an unexpected situation occurs around the vehicle, the first choice for action would naturally be or the vehicle to stop before colliding with any objects in front of it. If the self-driving car is unable to stop in time, the second response would be to swerve out of the way and avoid hitting any objects that would cause a crash, even if this means the car must go off the road. The ethical dilemma lies in the worst case scenario, where the self-driving car cannot stop in time nor can it avoid any collision. The issue is, if the vehicle has no option except crash into something, what should it hit? In addition to all the controversy over the proper course of action for a self-driving vehicle in such a situation, the software engineers working on the self-driving systems also must deal with the issue that they must decide on the proper course on behalf of society. They must then foist that decision upon the rest of society, even if a fraction of society ends up using self-driving cars.

Background:

As of May 5th of this year, Google self-driving cars have been reportedly involved in roughly 30 minor traffic incidents over the course of 5 million miles of road test, which compares closely to an estimated 21 traffic incidents for average human drivers [1]. It's important to keep in mind that of those 30 incidents, only two of them have been widely published as being the fault of the autonomous vehicle. The rest of them seemed to have been the human error of other drivers crashing into the self-driving cars. In the one incident, the self-driving car seems to have tried to merge into the same lane as a public bus, and the two vehicles hit their sides together because the autonomous algorithm assumed that the bus would slow down to let the car go first, but the bus did not slow down [2]. This seems to be the only incident that Google is willing to accept even a part of the responsibility for, because their algorithms and simulations forgot to account for the fact that larger vehicles are less likely to give way smaller cars because of their large inertia [2]. The test driver self-driving car apparently came to the same conclusion as the software, prior to the crash [2]. The other widely published incident came afterwards, but it too involved the software not knowing how to correctly perform a merge on a highway [3]. It seems that roads Google has not yet fully prepared their software for are considered forbidden routes in their database, and the self-driving algorithms will avoid using these roads. A self-driving car was on one of these forbidden highways and had an error when another sedan try to merge into traffic. There was no direct collision, But in order not crash the other car ended up spinning across the freeway into the grass while the self-driving car had to be taking over by a driver and ended up harshly swerving, which ended up injuring the passenger's spine [3]. Google denied causing an accident and completely shrugged off the responsibility for it by placing the blame on the other car, even though it was their software that did not behave politely on the road [3]. It's fairly clear that regulations for assigning responsibility for traffic incidents involving self-driving cars are severely lacking, and falling behind this new technology's progress.

So far, 30 states have passed at least one piece of legislation 2012 up until the present that deal with the definition, tasks performed, and permission to operate autonomous and self-driving vehicles. Naturally, these bills started in California in 2012 where the technology was first being developed, as well as the adjacent states that the vehicles would be most likely to cross into, but now states as far away as Florida and Maine now have bills concerning autonomous vehicles [4]. The legislation does evolve with time as the state learned more about autonomous vehicle technology, and becomes more comfortable with its use. In Florida's initial legislation in 2012, for example, it allows for the testing of autonomous vehicles under authorized conditions in authorized locations as long as the operator, who engages the self-driving technology, has a valid driver's license for that vehicle as well as insurance to take responsibility if something goes wrong. Florida revised its legislation in 2016, which allows autonomous vehicles on public roads removes the requirement that a person is actually in the vehicle, provided that the operator has a valid driver's license and the vehicles meet all the safety regulations of the road [4]. The more recent legislation in many states also allow for the testing and or use of platooning technology in the autonomous vehicles [4]. Platooning with self-driving vehicles is where several vehicles communicate with each other inform and electronically synchronized caravan for traveling at the same speed at a set distance apart. This allows for few, if any, drivers to have the vehicles tailgate one another in their slipstreams to save fuel at a distance apart that would be considered unsafe for humans to operate them at.

Discussion:

The most logical approach to the worst-case, 'there will be a crash' scenario is for the self-driving car to choose the path of collision that will kill the fewest people [5]. Preserving animal life is important too, but the majority of society values human life over pets or other animal life to such an extent that only people matter in this approach. This method applies to the classic trolley problem, and says that you should take the route to kill the one instead of the many, even if you end up killing several people in order to save many more. If the laws were written alone that line of thinking, it would indeed have its benefits in keeping of casualties to a minimum. Statistically speaking, this means that the occupants of the self-driving car would usually get the short end of the stick, especially in crowded cities and towns where there is typically more people in a crosswalk then can fit inside of a car. Thus, the big problem with that format of decision making in an autonomous vehicle is that much of the public would not want to ride in such a vehicle, due to the high tendency that their own car would sacrifice them for the greater good in the event of an unavoidable crash. That means the majority vehicles on the road would end up not being autonomous, thereby invalidating the entire point that self-driving cars would minimize human casualties, since very few would actually use them [5].

Since a consensus on this ethical dilemma would be almost impossible to achieve, one solution might be to offer multiple variations of the crash scenario decision algorithm, or have one version of the software in all of the self-driving cars that allows the user choose the settings on what outcome they prioritize the most. The good part about such a system would be the public would be rather happy with the freedom to choose how the software in their self-driving car is going to behave. Ideally, having these different preferences would not harm the overall system of an autonomous vehicular network in a significant way, because such circumstances that would call upon these preferences would be so few and far between that the roads would still be much safer than they are now. The reason for that, is if most of society converts over to driving

autonomous vehicles which interact rather well with each other in terms of avoiding collision, the human error that causes most traffic incidents would be almost entirely erased. However, if the manufacturer allows the user to acquire different versions or otherwise modify the moral algorithm, the user would then be responsible for any harmful consequences that their settings bring about [5].

In class, a very interesting point was brought up; the online resources all assumed that the occupants a self-driving car will automatically die in a crash. Fortunately, vehicles these days are loaded with all kinds of safety features, such as airbags and frame structures designed to self-sacrifice by absorbing all of the force of an impact, which crushes their structure and usually totals the car instead of allowing significant harm to come to the occupants. Another safety feature that a self-driving car has over a normal car, is that computers follow protocols to the letter. This means that they will not speed over the limit, and you would not be hit by another autonomous car going too fast, thereby greatly reducing the amount of impact force involved in these crash scenarios. After considering the probability of survival for each of the potential impact targets in a crash scenario, some of the opinions of what the software should have the car do shifted towards an increased effort to save the pedestrians.

Conclusion:

I too had been tripped up by the phrasing of the online articles, and thought that the death of the occupants of self-driving cars would be guaranteed. However, in light of the new perspective regarding the self-driving cars' built in safety measures, I think that the proper resolution would be a mix of the utilitarian 'save the most lives' approach and the 'my car is an armored shell' mentality. If I was going to be hit by a car, regardless if it was autonomous or not, I would rather be within another vehicle over being on the crosswalk, even if both vehicles were moving into a head-on collision. If I was in a self-driving car that had gone out of control, I believe the way to preserve the most life would be for my vehicle to swerve away from pedestrians, and collide with either a wall or parked vehicle on the side of the street. That does put all of the risk an impact force on the occupants of my vehicle, the reason behind consumers not wanting to buy the utilitarian autonomous software. However, the safety features of my vehicle would highly increase my chances of survival over those of some pedestrian whom on average tends to be rather...squishable. Therefore, I believe the autonomous vehicle should try to hit inanimate objects and try to not run over any people.

References:

[1]   *Waymo's Self-Driving Car Crash in Arizona Revives Tough Questions*, Aarian Marshall (https://www.wired.com/story/waymo-crash-self-driving-google-arizona/)

[2]   *Google Says It Bears "Some Responsibility" after Self-Driving Car Hit Bus*, David Shepardson (https://www.scientificamerican.com/article/google-says-it-bears-some-responsibility-after-self-driving-car-hit-bus/)

[3]   *GOOGLE SELF-DRIVING CAR CAUSED FREEWAY CRASH AFTER ENGINEER MODIFIED ITS SOFTWARE*, Jason Murdock (https://www.newsweek.com/former-google-engineer-allegedly-modified-self-driving-software-caused-crash-1174058)

[4]   *AUTONOMOUS VEHICLES | SELF-DRIVING VEHICLES ENACTED LEGISLATION*, NCSL (http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx)

[5]   *Why Self-Driving Cars Must Be Programmed to Kill*, MIT Technology Review (https://www.technologyreview.com/s/542626/why-self-driving-cars-must-be-programmed-to-kill/)