The Ethics of Driverless Cars

1. Introduction: Driverless cars are already here. Here is a brief timeline of events:

2009: Google begins testing its driverless car (Waymo) on public roads in California.
2016, Sept: Uber tests driverless fleet in Pittsburgh (then Toronto, Phoenix, and Boston).
2018, March: Driverless uber in Arizona kills jaywalking woman. Uber suspends program.
2018, May: Lyft tests driverless taxi fleet in Las Vegas.
2018, Oct: Google's Waymo clocks in its 10 millionth mile testing on public roads.
2018, Dec: Google Waymo tests driverless fleet in Phoenix.
2020: Tesla updates its cars with a fully-self-driving feature.
2020, Dec: GM tests driverless taxi fleet in San Francisco.
2023, Oct: Google Waymo launches truly driverless (no humans) robotaxi service in San Fran.
2023, Oct: GM fully driverless taxi hits and drags cyclist. GM suspends driverless program.

With the rise of driverless cars, suddenly everyone was talking about the moral implications (e.g., TED, TEDx talk, BBC, CBS, *Science*, etc.). Consider this case:

> **Trolley Revised:** You are in a driverless car, when suddenly five people run out in front of you. The only way to avoid hitting them is for the car to steer to the right. Unfortunately, there is one pedestrian to the right that the car will hit if it does so.

What should the car do?

If YOU were driving, you'd probably freak out and make a not-well-thought-out, "snap" decision in this case. But, with a driverless car we have the luxury of considering what to do, and programming the appropriate response perhaps YEARS in advance.

2. What Moral Principle? So, what should the car do? The instruction needs to be simple – something that can be applied to lots of different situations. How about this:

- **Do No Harm:** Never harm a human being.

Problem: In the original Trolley (Switch) case, it's a choice between doing harm (pulling the lever and killing 1) and allowing harm (doing nothing and letting 5 die). But that doesn't seem true of the autonomous car version. Arguably, the car is about to DO harm to someone NO MATTER WHAT. (*Imagine that you were driving. Could you really claim, after plowing into the five, that you merely allowed them to die?*)

[Also, perhaps it IS sometimes permissible to do harm to others. For instance, imagine that you had to run over someone's foot in order to get to a bus full of people in time to save them. In a choice between DO harm to one person's foot, or ALLOW 100 people to die, perhaps it's permissible to do the harm?]

• **Minimize Harm:** The car ought to do as little harm as possible.

Problem #1: Sometimes, this will require YOUR death. For example, imagine that the only way to avoid hitting the five people in the scenario above is for the car to steer off of a cliff, or a bridge, killing you inside of it. Are you okay with that?

Problem #2: Imagine that, in order to avoid a collision, your car can either steer left into a motorcyclist WITHOUT a helmet, or to the right into a motorcyclist WITH a helmet. Steering into the one who IS wearing a helmet will minimize harm. But, that makes it seem like you're "punishing" them for being more responsible. Is that okay?

Problem #3: Facial recognition technology is advancing quite quickly. In the near future, driverless cars will be able to identify pedestrians and other drivers instantly by sight (and perhaps also by pinging their devices to ID them). Imagine that the one person on the sidewalk in our example above is a famous heart surgeon; or imagine that they are very young, and that the 5 people in the street are very old. Wouldn't TRULY minimizing harm require saving the surgeon? Or the young person? Are you comfortable with autonomous vehicles having access to people's identities? Or with them assigning a score to the value of a life?

Problem #4: Often times, we think that it is justifiable to cause harm to someone when they are acting immorally. (For instance, this is why killing in self-defense is generally seen as morally acceptable, even if killing in general is not.) Now imagine that a driver and his friend decide to play "chicken" with a driverless car. The driverless car can avoid harming those two individuals by steering into one pedestrian instead. But, here, an innocent bystander is harmed in order to save two guilty wrong-doers. Should driverless cars be sensitive to questions about whether the surrounding individuals are doing something immoral, or even merely illegal?

Or, simply imagine that the 5 in the street in our main example are jaywalking. Do you STILL think that the car should kill the person walking legally on the sidewalk? If not, is jaywalking now a capital offense; the car now judge, jury, and executioner?

[Two more proposals, which we probably didn't discuss in class:

• **Legally-Adjusted Minimization:** The car ought to do as little harm as possible, adjusting for facts about whether the surrounding individuals are acting illegally.

Problem: But, we often draw a moral distinction between breaking the law accidentally versus with malicious intent. For instance, contrast a confused driver who accidentally drives up a one way street with a malicious driver who intentionally

does so for the purpose of putting people in danger. Driverless cars would not be able to distinguish between these two drivers' actions. Also: Sometimes a driver's behavior APPEARS illegal, but is actually legal. (For example, it is legal for a doctor to speed when rushing to the emergency room.)

• **Maximin:** On this proposal, it is not the TOTAL harm that would be minimized, but rather the WORST KIND of harm that would be minimized. For instance, it is clearly better to break two people's arms than kill one person. This is because death is a much worse KIND of harm.

Problem: But, then, we get into all kinds of sticky calculations. For instance, is paralyzing 100 children better than killing one other pedestrian?]

• **Total Passenger Safety:** The car ought to protect its own passenger(s) at all costs.

Problem: Imagine that the five in the cliff version of our main case are children. Better yet, imagine it's a BUS full of children, which will be pushed off of the cliff unless your car kills you instead. Are you really comfortable knowing that, in advance, it has already been determined that, given the option, you have already chosen to kill 50 children in order to save yourself?

[*Another complication: Should the lives of animals be factored into the car's consideration? For instance, should a car avoid hitting a dog, even if doing so requires that the car's paint job will be scraped by a guard rail?*]

• **Adjustable Ethics Settings:** Perhaps the consumer ought to be given several choices regarding which moral framework their car will adopt.

Problem: Plausibly, everyone will just select 'Total Passenger Safety', so this option just reduces to the previous one (though note that this would be a nice way for car manufacturers to absolve themselves of any blame, and places the responsibility solely on the customer instead).

Reply: We could pass regulations to set a "moral floor". Arguably, some actions are **morally obligatory**, while others are merely **supererogatory** (i.e., it would be a nice thing to do them, but morality does not require it). For instance, perhaps you would be obligated to rescue a drowning child if there is no one else around, but you are not obligated to send all of your savings to charity to rescue starving children. The latter would be a nice thing to do, but you are under no moral obligation to do it.

Still, the question remains: What would this moral floor look like? What is the minimum amount of sacrifice that morality requires of you?

Further Considerations:

- Driverless cars may be susceptible to hacking.
- They will put people out of jobs (truckers, delivery people, taxis, etc.).
- They may result in fewer total cars (you'd simply call driverless ubers when you want to go somewhere).
- They could spell the end of traffic jams (since the cars can synchronize).
- They will result in fewer kidney donations (due to fewer traffic fatalities).
- MIT is crowdsourcing driverless car ethics! Help them by taking the survey.