AK2003 Technology and Ethics

Who is morally responsible for fully autonomous cars?

Martin Barksten
barksten@kth.se

20th May 2015

1 Introduction

In 2012 one of Google's cofounders, Sergey Brin, in an interview says: "You'll ride in robot cars within 5 years". That is in just one year from now and while it might seem as if that is a bit too optimistic, robot cars, or self-driving cars as they are more often called, are not far from realistic. Google's self-driving cars have together driven more than 1.5 million miles.

The quick development of self-driving cars is driven by the many advantages that might be gained from getting humans off the road. In their annual report on traffic safety WHO points out that 1.2 million people die each year in traffic accidents [16]. And this is just looking at the traffic accidents that are fatal, low- and middle-income countries lose approximately 3% of their GDP as a result of traffic accidents. Studies also show that a majority of these accidents seem to be the cause "human error" [13], something an self-driving car would not be prone to do.

Further motivation to the gains from self-driving cars are described by Howard in [8], where he mentions the ability for disabled to use cars and the ability to revitalize failing cities.

But self-driving cars suffers from the problem of responsibility: who is to be held responsible when a self-driving car crashes? Or makes a seemingly odd decision? Some argue that the temporary solution is to allow the driver to intervene, temporarily taking control of the car to handle the situation. However, as Goodall points out this has several problems resulting in that the probability is high the driver will unable or unwilling to take control — making the car forced to take action anyway [4].

It therefore seems that the cars must be capable of taking all decisions themselves and that the problem of responsibility remain. This is a problem that needs solving, because as pointed out by Merchant et. al.: "Cars crash. So too will autonomous vehicles [. . .]" [10].

Neither is it hard to construct a scenario where a car will be forced to make an ethical decision. As an example let's use a variation of the commonly used Trolley problem outlined by Goodall in [3]: The car is driving on a small bridge and in the opposite lane is a bus, which suddenly turns towards the car's lane. The car then has three possible actions:

1. Drive off the bridge, guaranteeing a serious accident for the car;

2. Drive straight on towards a head-on collision with the bus, causing a less serious accident for both the car and bus;

3. Or attempt to drive past the bus with the possibility of a avoiding a crash, but a probability of much more serious injuries for the passengers of both the bus and the car.

It is not hard to imagine that a similar scenario will occur in reality, forcing the car to make a decision with moral implications — does it sacrifice the driver in favor of the bus which might contain more passengers? Does it try to save the driver? Or does it try avoid a collision, but at a might higher risk?

2 Essay question

This essay will attempt to find an answer to the following question: *Who is morally responsible for an autonomous car?* In order to do some distinctions have to be done.

First of all, when an entity is referred to as being morally responsible for an autonomous car it means that the the entity is responsible for the morally significant actions performed by the car. This definition follows the one defined by Eshleman in [2] who further illustrate what that entails by writing the following: "Thus, to be morally responsible for something, say an action, is to be worthy of a particular kind of reaction — praise, blame, or something akin to these — for having performed it."

When discussing the term autonomous car is used deliberately in order to describe a car that is fully autonomous, requiring nothing from any passenger in the car. The actions of the car are therefore decided by the car alone without any outside input.

The concept of moral responsibility will be split up into two more distinct terms: casually responsible and morally blameworthy. The former, casually responsible, refers to being the actor that takes responsibility for the actions of the autonomous car. The latter, morally blameworthy, means being to blame — in moral terms — for the actions of the autonomous car. To illustrate the difference between the two, consider a scenario where the car crashes and causes damage to a property to save five persons' lives. The manufacturer might be considered casually responsible for the action, but seeing as the car saved five lives they are not be morally blamed for the action taken by the car.

The car is considered not to be an moral agents; without specifying the criteria for what it means to be a moral agent it seems unlikely that the autonomous cars coming out in the coming years will fulfill even the most basic of criteria. As such, the possibility is not taken into account in the analysis.

Finally, the introduction of autonomous cars is considered to be good and something to strive for. It is assumed that traffic accidents will decrease and that there will be a positive gain from introducing the cars. This is not yet proven to be true [11], but discussing the case where they are worse than human drivers is considerably less interesting (see [10] for further motivation).

3 Analysis

In order to analyze who is morally responsible for an autonomous car I will go through all actors that might be potentially considered responsible — discussing what the consequences of that actor being responsible would mean in terms of duties etc.

The actors that might be considered responsible for an autonomous car are:

1. The passenger or passengers of the vehicle;

2. The manufacturer responsible for programming, constructing and selling the car;

3. And the government that set the legal framework allowing the vehicles.

For each of these actors an analysis of the moral responsibility for that actor in two different scenarios will be discussed. The two scenarios are described below:

1. *The avoidable scenario* — where the vehicle has the potential to cause an accident, but the possibility of avoiding it if acting correctly;

2. *The unavoidable scenario* — where the vehicle must choose between several different wrong possible actions — all of which will yield an accident of some sort.

Scenario 2 might be considered an example of the Trolley problem, a scenario where an actor must choose between several outcomes none of which is the obviously correct or incorrect one [14].

The analysis will be done from two perspectives:

- A deontological perspective based on the idea that "ought implies can", which extends to the fact that moral responsibility derives from an ability to affect the outcome (see [12][Ch. 7] for a more in-depth discussion);

- And a consequentialistic perspective where the focus is on what the positive consequences will be of holding the actor morally responsible.

The following three sections will therefore analyze each actor in turn, considering the two scenarios described above.

3.1 The passenger

For the sake of this analysis it is considered to be only one passenger in the car; the difference between one or many passengers is of minor relevance to the analysis.

For the sake of this discussion it is important to consider what decisions the passenger made that might be considered to influence the outcome in the two scenarios. The most obvious decision was to choose to use an autonomous car, had the passenger not used the car none of the two scenarios would have happened.

However, a less obvious decision made by the passenger is also that of choosing car. If, as seems likely, there are several competing cars with different implementations and thus different ways of acting — the passenger has made a deliberate decision regarding the actions that the car will take in traffic.

3.1.1 The avoidable scenario

From a deontological perspective the passenger can be held morally blameworthy for those actions that they could affect. As mentioned above, these are choosing to use an autonomous car and choosing which car to buy and use.

Hevelke et. al. point to the fact that using a car will lead to an increased risk of an accident, thus making the passenger to at least a small degree blameworthy [7]. From this they conclude that the passenger is to a small degree blameworthy, although it is not the individual passenger's fault — it was just bad luck the accident occurred — but rather all user's of autonomous cars.

Another thing to take into account is if the passenger deliberately chose a cheaper and less safe car in which case the passenger could affect the outcome to some extent, adding blameworthiness to them.

Holding the passenger responsible for the actions of the car would potentially have positive effects in reducing car usage, while not restricting it — hopefully reducing traffic accidents. As such, from both points of views the passenger is — at least to a small degree — morally blameworthy.

3.1.2 The unavoidable scenario

In the case where the car can not avoid a crash and instead must make a moral decision the blameworthiness of the outcome can to some extent be considered to be the passenger's. If the passenger was presented two cars, one implementing Consequentialism and one Deontology, the moral action taken by the car is just an extension of the passenger's choice of morality.

Once again, it is important to point out that if the passenger chose between two cars without knowing the morality of either, the consequences were out of the passenger's control and thus they can not be held morally blameworthy.

For a consequentialist in the case where the scenario is unavoidable holding the passenger responsible would lead to an increased responsibility when buying a car and choosing how it should act. Given that sufficient information exists so that the passenger can make an educated decision, this might improve the passenger's feeling of responsibility and awareness of the issue.

3.2 The manufacturer

The manufacturer is considered as one single actor, no distinction is made between individuals in the company or other manufacturer's that provide parts to manufacturer.

The manufacturer is the actor taking the most actions of all actors involved as it constructs, designs and programs the car thus being in control of almost every aspect of it.

3.2.1 The avoidable scenario

Intuition would probably tell us that the manufacturer is to blame for an autonomous car that makes a mistake seeing as they are the ones to built the car and thus the ones to introduce the code leading to the faulty decision process. This is an argument that relies on the deontological way of thinking that the manufacturer's were those that could affect the outcome, thus those that ought to have fixed the problem.

However, as Goodall says: "Any system ever engineered has occasionally failed." [4]. It is therefore rather a question of whether the company chose to not be thorough, thus putting the blame on them. Or if they were as thorough as might be considered reasonable, in which case they are not morally blameworthy as they did what they ought to do.

Once again the Ford Pinto case (mentioned in Section 3.1.1) is similar in that disregarding safety without informating makes the company morally blameworthy. If the company informs the public, they shift the moral blameworthiness to the passenger.

This conclusion agrees with what a consequentialistic view might arrive at as well: companies disregarding safety should be held morally blameworthy for doing so, but if they do *the best possible* holding them morally blameworthy would discourage further development of the cars.

3.2.2 The unavoidable scenario

From a deontological perspective it seems hard to argue that the manufacturer is to blame — they did what they could. This is also supported the fact that holding manufacturers responsible may deter them from developing or increase the prices of the cars, neither of which is advantageous [10].

Once again a consequentialist point of view would agree that holding the manufacturer responsible would only discourage development of autonomous cars, while not providing seemingly very many positive benefits.

3.3 The government

The government refers to the authority that allowed the autonomous car on the road and set the legal framework for what is considered a legal autonomous car.

Once again we need to establish what decisions the government made; they allowed autonomous cars on the road and they defined what criteria an autonomous car need to fulfill in order for it to be allowed.

3.3.1 The avoidable scenario

In this scenario it seems that the moral blameworthiness can be put on the government if the autonomous car is found to be legal, yet still acts incorrectly. In that case, the government could reasonably have done more, and thus ought to have done so. Thus, they are in that case at least partially morally blameworthy.

If the car on the other hand was illegal it becomes a matter of enforcing the law — did the government do what they could to ensure that all cars on the road abide the law? If they do not, the government has at least partial moral blameworthiness.

This also agrees with that a consequentialist would argue; if we do not hold the government blameworthy to some extent then the companies will continue to build bad cars, causing more accidents and deaths.

3.3.2 The unavoidable scenario

In the case where the crash is unavoidable some blame can be put on the government as they allowed the car on the road, by extension allowing the accident to happen. However, consider the cars will save lives and reduce the number of accidents, so in that regard the government did what they ought to — they allowed autonomous cars. The moral blameworthiness therefore seems hard to put on the government in this case as they did what they ought to do — which is to try to prevent lives.

Additionally there is little to gain from holding the government morally blameworthy, the only real consequence it could have would be that the cars become illegal and that would mean an increase in accidents and deaths.

4 Summary

To summarize the analysis given above we will once again return to the two scenarios described.

4.1 The avoidable scenario

In this scenario I have argued that the passenger is to be held partially blameworthy if it is the case that they have made an informed decision in selecting the car to use.

The manufacturer is also to be considered blameworthy if it the case that they disregarded safety, or failed to inform the consumer of the risks involved in using their autonomous car. If they were what might be considered reasonably thorough in the production of the car, they are not be held morally blameworthy.

Finally, the government is to be considered morally blameworthy if it is the case that the car was legal and acted incorrectly, or illegal and the law insufficiently enforced.

A result of this is that no actor could be held morally blameworthy if it is a the government finds the car legal and ensures that is the case, the manufacturer gives sufficient regard for safety and are thorough in the production of the car and finally the passenger's decision to buy a car was made with no information that might be considered morally relevant. Of course, this all seems rather unlikely.

4.2 The unavoidable scenario

In this scenario the passenger is morally blameworthy if it was the passenger's informed choice of car that dictated the action taken by the car.

I argued that the manufacturer is not to be held morally blameworthy in this scenario as they did what they could and holding them responsible would have minor positive consequences.

Finally, I argued that the government acted correctly given the basic assumption underlying this discussion — the car's will reduce accidents and save lives. Apart from this there is also once again little to gain from holding the government responsible.

As a result, it is likely that no actor can be held morally blameworthy in this scenario. Something that seems to agree with an intuition that in a scenario with a negative outcome, but one every actor did their best to prevent — no one is to be blamed.

References

[1] *Average Annual Miles per Driver by Age Group.* URL: `http://www.fhwa.dot.gov/ohim/onh00/bar8.htm` (visited on 05/10/2016).

[2] Andrew Eshleman. "Moral Responsibility". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Summer 2014. 2014. URL: `http://plato.stanford.edu/archives/sum2014/entries/moral-responsibility/` (visited on 05/11/2016).

[3] Noah Goodall. "Ethical Decision Making During Automated Vehicle Crashes". In: *Transportation Research Record: Journal of the Transportation Research Board* 2424 (Dec. 2014), pp. 58–65. ISSN: 0361-1981. DOI: `10.3141/2424-07`. URL: `http://trrjournalonline.trb.org/doi/10.3141/2424-07` (visited on 05/10/2016).

[4] Noah J. Goodall. "Machine Ethics and Automated Vehicles". In: *Road Vehicle Automation.* Ed. by Gereon Meyer and Sven Beiker. Cham: Springer International Publishing, 2014, pp. 93–102. ISBN: 978-3-319-05989-1 978-3-319-05990-7. URL: `http://link.springer.com/10.1007/978-3-319-05990-7_9` (visited on 05/10/2016).

[5] *Google Self-Driving Car Project.* Google Self-Driving Car Project. URL: `http://www.google.com/selfdrivingcar` (visited on 05/10/2016).

[6] *Google's Sergey Brin: You'll ride in robot cars within 5 years - CNET.* URL: `http://www.cnet.com/news/googles-sergey-brin-youll-ride-in-robot-cars-within-5-years/` (visited on 05/10/2016).

[7] Alexander Hevelke and Julian Nida-Rümelin. "Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis". In: *Science and Engineering Ethics* 21.3 (June 11, 2014), pp. 619–630. ISSN: 1353-3452, 1471-5546. DOI: `10.1007/s11948-014-9565-5`. URL: `http://link.springer.com.focus.lib.kth.se/article/10.1007/s11948-014-9565-5` (visited on 05/09/2016).

[8] Don Howard. *Science Matters » Blog Archive » Robots on the Road: The Moral Imperative of the Driverless Car.* URL: http://donhoward-blog.nd.edu/2013/11/07/robots-on-the-road-the-moral-imperative-of-the-driverless-car/#.VzHclBV96bX (visited on 05/10/2016).

[9] *Licensed Drivers - Our Nation's Highways - 2000.* URL: http://www.fhwa.dot.gov/ohim/onh00/onh2p4.htm (visited on 05/10/2016).

[10] Gary E. Marchant and Rachel A. Lindor. "Coming Collision between Autonomous Vehicles and the Liability System, The". In: *Santa Clara L. Rev.* 52 (2012), p. 1321. URL: http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/saclr52&section=41 (visited on 05/10/2016).

[11] Brandon Schoettle and Michael Sivak. "A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles". In: (2015).

[12] Robert Stern. *Kantian Ethics: Value, Agency, and Obligation.* Oxford University Press, Oct. 1, 2015. ISBN: 978-0-19-872229-8. URL: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198722298.001.0001/acprof-9780198722298 (visited on 05/17/2016).

[13] *The Relative Frequency of Unsafe Driving Acts: Summary.* URL: http://www.nhtsa.gov/people/injury/research/udashortrpt/summary.html (visited on 05/11/2016).

[14] *Trolley problem.* In: *Wikipedia, the free encyclopedia.* Page Version ID: 719408859. May 9, 2016. URL: https://en.wikipedia.org/w/index.php?title=Trolley_problem&oldid=719408859 (visited on 05/10/2016).

[15] Ibo van de Poel and Lambèr Royakkers. *Ethics, Technology, and Engineering : An Introduction.* 1st ed. Chicester: Wiley, 2011. ISBN: 978-1-4443-9570-9.

[16] World Health Organization. *Global status report on road safety 2015: supporting a decade of action.* Geneva, Switzerland: WHO, 2015. ISBN: 978-92-4-156506-6.