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, states in an interview: "You'll ride in robot cars within 5 years". That is 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 very far away. Just Google's own self-driving cars have together driven more than 1.5 million miles [5].

Self-driving cars have seen a quick development in recent years, a development driven by the many advantages the cars offer, the primary of which might be reducing traffic accidents. Several studies show that a majority of traffic accidents seem to be the cause of "human error" [13], accidents self-driving cars could avoid. Add to that, that 1.2 million people die each year in traffic accidents [17] and that they cost us considerably — low- and middle-income countries lose approximately 3% of their GDP as a result of traffic accidents — and you should have sufficient motivation to why self-driving cars would be beneficial.

There are also many more advantages to be gained from self-driving cars and Howard describe two important ones in [8], where he mentions the ability for disabled to use cars and the ability to revitalize failing cities.

However, self-driving cars suffer from the problem of responsibility: who is to be held responsible when a self-driving car crashes? Or when it makes a seemingly odd decision? One argument is that a temporary solution is to allow the driver to intervene, temporarily taking control of the car to handle the situation. However, as pointed out in [4], this is problematic and it is likely not realistic.

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

It is not hard to imagine a scenario where a car will be forced to make an ethical decision. As an example let's use an example scenario 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 is presented with three possible actions:

1. Drive off the bridge, guaranteeing a serious accident for the car;
2. Drive straight towards a head-on collision with the bus which would cause a less serious accident for both the car and bus;
3. Or attempt to drive past the bus with the possibility of 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 similar scenarios will occur in reality, forcing the car to make decisions with moral implications — does it sacrifice the driver of the passengers of the bus? Does it try to save the driver? Or does it try to avoid a collision, but at a higher risk?

2 Essay question

The aim of this essay is to find an answer to the following question: *Who is morally responsible for an autonomous car?* In order to do so, 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 the term responsibility is used it is always according to the definition above and none of many other possible interpretations (see for example [16, Ch.2]). It is especially important to note that outcome responsibility — i.e. being responsible for the outcome of a scenario — does not necessarily imply moral responsibility. A manufacturer might be considered responsible for the outcome of a crash involving one of its self-driving cars, but that does not necessarily mean that they are also morally responsible.

Throughout the report the term autonomous car is used to describe a fully autonomous car which requires nothing from the passenger at all. The actions of the car are therefore the consequence of decisions made by the car without the possibility for any external input.

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

The analysis will be done from two different points of view:

- 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 based the idea that the actor who should be held morally responsible is the one for which doing so will provide the best consequences.

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 from the two perspectives outlined.

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 setting 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 crashes, but had the possibility of avoiding it if it had acted correctly;
2. *The unavoidable scenario* — where the vehicle crashes and had to choose between several different possible actions — all of which would have yielded an accident of some sort (consider it as a variant of the famous Trolley problem [14])

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

3.1 The passenger

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

The actions the passenger made that might be considered to have influenced the outcome in the scenarios are:

- Choosing to use an autonomous car, had the passenger not done so there would have been no possibility for either of the two scenarios to occur;
- And choosing which car to use between — as seems likely to be the case — several competing models.

3.1.1 The avoidable scenario

The action of choosing to use an autonomous car is one that greatly affects the outcome as if the passenger had not chosen to do so, no crash would have occurred. It therefore seems that the passenger should be held to some degree blameworthy for the crash as they could affect the outcome. However, as Hevelke et. al. point out, it is not the individual passenger's fault that the car crashed, they did not do anything worse than any other passenger in an autonomous car [7]. As a consequence it is not the individual passenger that is to be held

blameworthy, but rather all user's of autonomous cars that are to be held to some degree blameworthy for all autonomous cars.

The second action, choosing which car between several competing models also is one that could bear with it some moral blameworthiness. Consider that the user had the ability to buy a more expensive and safe car but chose the less safe one involved in the crash. In this case, the passenger could clearly affect the outcome to some extent, adding blameworthiness to them.

From a consequentialistic point of view it seems that holding the passenger to some degree responsible for the actions of the car could have possible consequences in reducing car usage in favor of other methods of transportation. Compared to the current responsibility held by the driver, who is almost fully responsible for a car, it would still mean a reduced responsibility and a motivation to use an autonomous car rather than a normal one.

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 that could be said to "perform the majority of the actions" out of the three actors considered, the reason is that the manufacturer designs, constructs and programs the car — putting them in control of almost every aspect of it.

3.2.1 The avoidable scenario

It seems intuitive that the manufacturer should in the avoidable scenario be held responsible, they are the ones that built the car and thus the ones that introduced — or didn't fix — the error that caused the car to not avoid the crash.

However, as Goodall points out: "Any system ever engineered has occasionally failed." [4]. It should therefore rather be discussed whether the company were sufficiently thorough in regards to safety. If they are found to be *reasonably thorough* they did what they could and as such are not morally blameworthy.

A good example of this might be the Ford Pinto case [15, Ch.3] where Ford could have avoided a three people dying in a car crash if they had spent $11 more dollars. It seems that they reasonably could have done more to prevent the accident and thus ought to have done so.

This does not necessarily mean the manufacturer must make their car as expensive as possible to introduce all safety concerns, they might also inform the consumers of the safety risks and allow them to make the choice instead. In doing so the manufacturer respects the autonomy of the consumer and shifts moral blameworthiness to them instead (as also discussed in Section 3.1)

This conclusion agrees with what a consequentialistic view might arrive at as well: companies disregarding safety should be held morally blameworthy for doing so as it would force them to improve the safety of the cars, thus reducing accidents. If a company however are found to have done *the best possible*, holding them morally blameworthy would only discourage manufacturers from making autonomous cars — which would have negative consequences.

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.

From a consequentialistic point of view holding the manufacturer responsible even though they did what they could would improve nothing and only deter them from making autonomous cars [10].

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.

We need to establish what actions has made that influenced the outcome:

- They defined what standards an autonomous car need to uphold to be allowed on the roads;
- And they chose to which extent these standards are controlled and enforced.

3.3.1 The avoidable scenario

If the autonomous car acts incorrectly even though it is legal it seems as if the government could reasonably have done more by setting stricter laws. Thus, they are 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 have reasonably done to ensure

that all cars on the road abide the law? If they did not, the government has at least partial moral blameworthiness.

If we assume the companies to act only according to the law and not take morality into consideration, it must the government's job to ensure that the law and morality are as close as possible. Thus, a consequentialistic point of view agree that the government need to be held responsible if the laws or the enforcement of them is insufficient — otherwise, the cars will see no improvement.

3.3.2 The unavoidable scenario

In the case where the crash is unavoidable it could be argued that some blame could be put on the government, when allowing the cars they by extension allowed the accident to happen. However, the basis for this discussion is that the cars are better than human drivers. This means that the government did what they could to reduce traffic accidents, they allowed the autonomous cars. As they did what they ought to do, they are not to be held morally blameworthy.

Neither does there seem to be any positive consequences of holding the government responsible in this scenario. The only consequence that seems to come from holding them responsible is complete ban of all cars — which is clearly not positive.

4 Summary

To summarize the analysis given above I will once again return to the two scenarios and summarize how the blame is divided between the actors. Finally, I give a short conclusion that should answer the question posed at the start of this essay: *Who is morally responsible for an autonomous car?*

4.1 The avoidable scenario

In this scenario I have argued that all user's of autonomous cars are to be held morally responsible for the actions of the cars. Furthermore, if the passenger has made an informed decision when selecting the car to use they also take some moral responsibility.

The manufacturer is blameworthy if disregarded safety or failed to inform the consumer of the risks involved in using their autonomous car. Otherwise, they are not morally blameworthy.

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

As a result, the passenger will always be to some degree morally blameworthy for the actions of the car whereas the manufacturer and government can avoid moral blame if they are able to provide proof that they have done what could have bean reasonably asked of them.

4.2 The unavoidable scenario

In this scenario the passenger is to a small degree morally responsible for the car if the passengers choice of car was deliberate and informed and included deciding how the car should act morally in an unavoidable scenario.

I argued that the manufacturer is not to be held morally blameworthy in this scenario as they did what they could and it would provide nothing but negative consequences to hold them responsible.

Finally, I argued that the government acted correctly given the basic assumption underlying this discussion — that the cars will reduce accidents and save lives. Apart from this there also seem to be no positive consequences to holding the government responsible.

As a result, the moral responsibility for the car in this scenario can only to a small degree be put on the passenger. The manufacturer and government are neither morally blameworthy. This seems to some extent intuitive; in a scenario with a negative outcome, which every actor did their best to prevent, no one is to be blamed.

4.3 Conclusion

To conclude, who is morally responsible for an autonomous car depends on the actions taken by the actors involved and the responsibility is usually split between one or more of them.

For the avoidable scenario, the moral responsibility is split between all actors involved, with the possibility of the manufacturer and government to motivate that they performed their duties to a sufficient degree as to not be held morally blameworthy.

For the unavoidable scenario the responsibility can be put on the passenger if the choice of car was also a choice of how a car should act morally. The manufacturer and government are found to not be morally responsible.

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*. 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] Nicole A. Vincent, Ibo van de Poel, and Jeroen van den Hoven, eds. *Moral Responsibility*. Vol. 27. Library of Ethics and Applied Philosophy. Dordrecht: Springer Netherlands, 2011. ISBN: 978-94-007-1877-7 978-94-007-1878-4. URL: http://link.springer.com/10.1007/978-94-007-1878-4 (visited on 05/09/2016).

[17] 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.