Ethics of autonomous cars

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Group for Ethics in Information Technology
Seminar Business, Technology and Ethics

Vorgelegt von:	Tim Jammer
E-Mail-Adresse:	3jammer@informatik.uni-hamburg.de
Matrikelnummer:	6527284
Studiengang:	Informatik
Betreuer:	Dr. Pak Hag Wong

Hamburg, den March 27, 2018

Abstract

Within this paper I want to discuss some points on wether or not the usage of autonomous vehicles can be morally responsible.

Therefore in Section 1.1 I will give a short introduction and definition of "autonomy". Section 1.1.2 follow up with an overview over the main benefits the adoption of automated driving might yield.

In Section 1.2 I want to discuss questions with regards on how an autonomous car should crash in a trolley problem like situation more in detail (Section 1.2.2). Furthermore I want to address the restriction of personal freedom (Section 1.2.1) that might occur when the car takes over the control in these situations.

Section 1.3 describes very briefly some other related questions that should also be considered.

I will conclude with my final thoughts in Section 1.4.

1 Ethics of autonomous cars

1.1 Introduction

Whithin this paper I want to address the question whether or not automous driving of cars is a moral technology. Therefore I will give a briefly overview of this technology.

1.1.1 Automation leves

The National Highway Traffic Safety Administration of the United States defines five levels of automation: [A$^+$13]

Level 0 **No-Automation**: The driver is in complete and sole control of the primary vehicle controls (brake, steering, throttle, and motive power) at all times, and is solely responsible for monitoring the roadway and for safe operation of all vehicle controls. Vehicles that have certain driver support/convenience systems but do not have control authority over steering, braking, or throttle would still be considered “level 0” vehicles.

Level 1 **Function-specific Automation**: Automation at this level involves one or more specific control functions; if multiple functions are automated, they operate independently from each other. The driver has overall control, and is solely responsible for safe operation, but can choose to cede limited authority over a primary control (as in adaptive cruise control), the vehicle can automatically assume limited authority over a primary control (as in electronic stability control), or the automated system can provide added control to aid the driver in certain normal driving or crash-imminent situations (e.g., dynamic brake support in emergencies).

Level 2 **Combined Function Automation**: This level involves automation of at least two primary control functions designed to work in unison to relieve the driver of control of those functions. Vehicles at this level of automation can utilize shared authority when the driver cedes active primary control in certain limited driving situations. The driver is still responsible for monitoring the roadway and safe operation and is expected to be available for control at all times and on short notice. The system can relinquish control with no advance warning and the driver must be ready to control the vehicle safely. An example of combined functions enabling a Level 2 system is adaptive cruise control in combination with lane centering. The major distinction between level 1 and level 2 is that, at level 2 in the specific operating conditions for which the system is designed, an automated operating mode is enabled such that the driver is disengaged from physically operating the vehicle by having his or her hands off the steering wheel AND foot off pedal at the same time.

Level 3 **Limited Self-Driving Automation**: Vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions and in those conditions to rely heavily on the vehicle to monitor for changes in those conditions requiring transition back to driver control. The driver is expected to be available for occasional control, but with sufficiently comfortable transition time. The vehicle is designed to ensure safe operation during the automated driving mode. An example would be an automated or self-driving car that can determine when the system is no longer able to support automation, such as from an oncoming construction area, and then signals to the driver to reengage in the driving task, providing the driver with an appropriate amount of transition time to safely regain manual control. The major distinction between level 2 and level 3 is that at level 3, the vehicle is designed so that the driver is not expected to constantly monitor the roadway while driving.

Level 4 **Full Self-Driving Automation**: The vehicle is designed to perform all safety-critical driving functions and monitor roadway conditions for an entire trip. Such a design anticipates that the driver will provide destination or navigation input, but is not expected to be available for control at any time during the trip. This includes both occupied and unoccupied vehicles. By design, safe operation rests solely on the automated vehicle system.

There already exist vehicles for all automation levels. For example the Deutsche Bahn already operates level 4 automated busses. [AG17]

Within the following discussion I will focus on vehicles of automation level 4 as defined above. Although most of the points will also apply to automation level 3 .

1.1.2 Benifits of automated driving

I think that there will be many benefits of fully automated driving.

First, fully automated driving will allow elderly or disabled people to stay mobile. This is especially important in rural areas where no sufficient public transport is available.

Additionaly Figure 1.1 shows the amount of traffic accidents in Germany in the year 2016 grouped by the course of the accident. One can see that many accident relate to some fault of at least one involved driver. I assume that autonomous cars will not make that many faults. For example accidents which are caused of some missunderstanding of the rules for right of way might even be entirly eliminated due to car to car communication, where the cars will agree who has the right of way and in case of some uncertainty they

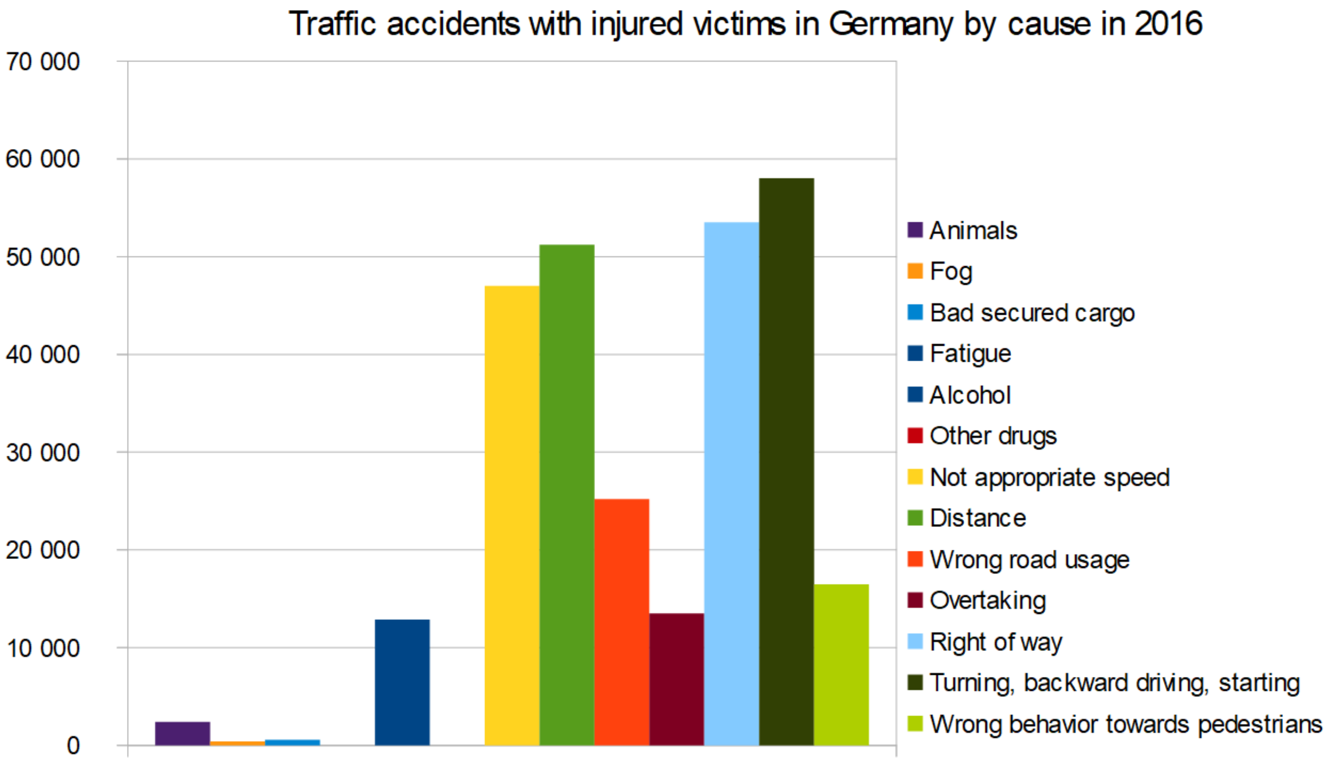

Figure 1.1: Traffic accidents with injured victims in Germany. Data from [Bun]

will just stop. If the technology of automated driving can save lives due to less accidents it implies to me it is morally acceptable to use them. If they are much safer then human drivers it should also be our moral responsibility to use autonomous driving.

Of cause with the technology of automated driving many moral questions arises. Within the next sections I want to discuss some of them.

1.2 Ethic crashing

One of the most discussed question is how should an autonomous car behave in a trolley problem.
In its original form a trolley problem is a theoretical thaught experiment where a train is heading towards five people on a track and can not stop in time. You are standing some distance off in the train yard, next to a lever. If you pull this lever, the trolley will switch to a different set of tracks. However, you notice that there is one person tied up on the side track. The question is whether or not you pull the lever and save the five people but kill the person on the other track. The problem is visualized in Figure 1.2.

The problem can also be easily transferred to cars not driving on fixed tracks. For example imagine you drive on a road with a sidewalk right side. Suddenly and not

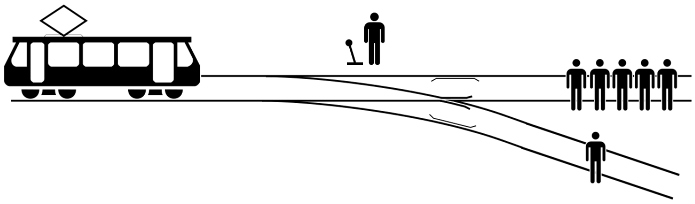

Figure 1.2: An illustration of the trolley problem.
from `https://commons.wikimedia.org/wiki/File:Trolley_Problem.svg`

foreseeable the truck on the left lane driving in the other direction start skidding. You are only left with two options: Crashing into the truck which will lead to your death or evade the truck but thereforerun over the pedestrians on the sidewalk to your right.[1]

One important point to consider is that for an autonomous car (respectively for its programmers) this problem is not only a theoretical thought experiment anymore. An autonomous car is capable of taking split second decisions. Therefore this problem need to be addressed even if situations like this will only occur very, very rarely.

Within the next sections I want to discuss two related questions to the trolley problem and the decision an automated vehicle should take.

1.2.1 Restriction of freedom

First there is a restriction of freedom if an automated vehicle takes moral decisions over your life because one is not in change of this decision himself anymore.

Some might argue that this limitation of the personal freedom of a human is not acceptable at all. It can therefore not be a moral responsibility to use autonomous cars.

I think that this is not the case. It is already a moral responsibility to only drive a car when one is sufficiently capable of doing so. If one is not sufficiently capable of driving (i.e. because he is drunk) it already is his moral obligation to give away that decision over his life to others.[2] For example if one chooses to use a taxi because he can not drive himself anymore, than he is froced to give away the morle decision over his own life in a trolley problem like situation to the taxi driver.
Therefore I want to ask the question where is the difference in giving away that decision to a taxi driver, a random stranger one just met, as opposed to giving away that decision to a machine respectively to its programmers, who are also random strangers one do not know?

[1] I choose this example because this situation can arise even if the car driver has made no fault at all (for example driving to fast) which for me is an essential part of a trolley problem.

[2] Here he gives away some of his freedom (not driving himself) to protect the freedom of others (not getting overrun by an uncapable driver)

In either way one has to assume that theese individuals will implement this decision the best of their knowledge and belief.

1.2.2 Utitlitarian cars

If one wishes to accept the restriction of freedom this still does not answer the question on how an autonomous vehicle should behave in a trolley problem like scenario.

The first question for me is, why do we not just trust the programming team that they will hold moral responsibility as above, analogous to a taxi or bus driver today. One important point is that a taxi or bus driver takes this kind of decision for all inmates of the vehicle including himself whereas the programmers of an autonomous car are not directly involved in the situation to juge. Furthermore they have to come up with a sort of overall guidelines for the vehicle as they cannot forsee any situation in the future. Therefore the society should agree on a set of overall guidlines that a programming team can follow. In Germany there is an ethics commision of the German Federal Ministry of Transport and Digital Infrastructure for this pupuse.
It is important that the society agrees on an overall set of rules, because it will result in a prisinors dilemma if every car owner can choose the guidlines for the car in a trolley problem himslef. [GM17]

A common concept states that autonomous vehicles should act in an utilitarian way and minimize harm. In a survey [BSR15] found out that most participants agree to such guidelines. But if the participants should imagine themselvs in the vehilce involved the live of the vehilces occupants should be valued more, while still following general utilitarian guilelines.

One problem with utilitarian guidlines might be that they will contain a bias against people who value safty more. Imagine a trolley problem where an autonomous vehicle can choose whether to crash into a cyclist with a helmet or into a cyclist without the helmet. Under a utilitarian guideline it will choose the option which introduces less harm. Then it would collide with the cyclist with the helmet because he would suffer less severe injuries as he has a helmet protecting him. One might argue that this is an unfair behaviour and the car should crash into the cyclist without the helmet in order to "reward" the cyclist with a helmet that he was paying more attention to safty. Of cause the cyclist without a helmet have had the option to take a helmet but he was willingly taking the risk of getting injured severly and therfore it is not moraly inresponsible to inflict more damege to him.

Additional the ethics commision of the German Federal Ministry of Transport and Digital Infrastructure stated that any jugement based on personal traits like age, gender, physical or mental constitution is unethical. [fVudI17]
I agree with this statment. One implication of this is that any kind of general rule on how to behave in a trolley problem is in itself unethical. Because it weights lives of

differnt humans against one another which is not acceptable. In some circumstantes such decisions might not be avoided but a general strategy on how to deal with this situations is not appropriate. Rather it should be a case by case decision. Either option in a trolley Problem is the false option because it involves the death of people. One cannot take the "right" decision. For me it is an essential part of freedom to accept that different individuals have different opinions on the "right" decision in a trolley problem. Therefore I have to accept the consequences of other people taking the option, which I consider more false than the other availible option, whithout seeing any form of a morally irresponsible behaviour of the people.

For me this is not contradictory to my point of view that autonomous vecicles are an ethical technology. As [fVudI17] states, the first and foremost priority in the development of autnomous cars should be to minmize the risk that a trolley problem like situation will become reality in the first place. As above an morraly irresponsible decision is not taken in a trolley problem like situation. The irresponsibel decision was taken before the trolley problem like situation arises and leads to the dangerous situation[3].

As long as the programming team will do anything possible to ensure a trolley problem like scenario can never occur i will accept the programmers decision on how the car should behave in a trolley problem as an expression of the programmers personal freedom. Just like I do now with taxi, bus or other human drivers.

1.3 Other moral questions to ask/judge

It is important to remember that the question on how should an autonomous car crash is not the only important question to consider. In this section I want to briefly discuss some other issues with the usage of automated driving.

1.3.1 Jevons paradox

William Stanley Jevons (01.09.1835 to 13.08.1882) stated that if technological progress increases the efficiency of usage of a ressource the total usage of that ressource will increase overall. This is due to the fact that increased efficienty leads to lower prices. This causes more demand because more people are willingly to use this ressource more often as they can now afford this.[Alc05]

This paradox might also apply to the usage of autonomated driving. Especially for the transportation industry automated driving will decrease the cost of transporting things around. This might lead to more goods being transported which causes more cars to drive around. With more cars there might be more accidents in total.
I think that this will not be a moral issue even if there might be more accidents in total. If all the trucks drive without a driver many accidents will happen without any human

[3] e.g. driving too fast

getting harmed as humans will not be present in most vehicles and all vehicles can be programmed to always choose a vehicle without a human on board to collide with.

1.3.2 Automation and work

Another issue might be that almost all humen drivers will loose their job as the machine can drive the car and -this counts especially for the logistigs sector- there is no need for a human to stay on board.
I do not want to discuss this in detail as this is not only an important topic to discuss in relation to autonomous cars. Currently many other jobs are automated as well. [MCM+17] Therefore this problem is an important topic to discuss in the society right now and not an newly emerging question related to the usage of autonomous driving.

There are still many more important points we need to consider, for example:

- "many organ transplants come from car-accident victims, how will society manage a declining and already insufficient supply of donated organs?" [Lin14]
- Autonomous cars can be used to oversee or control the freedom of movement.[4]
- Automated vehicles may be vulnerable to cyberattacks. Especially frightening are adversarial learning techniques to attack a machine learned behaviour because there are no means of protection against this kind of attack. [Jar17]

But this paper can only present a very short overview over all these numerous other questions. Personally I think that the answer to many of the question stated above lies in other new or improved technology. Nevertheless it is important to keep these questions in mind while further developing autonomous cars.

[4] Surveillance of movement is already possible due to mobile phones. It is enough to have the telephone number to kill someone. [Fed16] Therefore this is not an newly arising issue of autonomous cars.

1.4 Conclusion

I think that autonomous driving is an ethical technology.

I can not see any newly arising moral issues with the usage of autonomous cars. All important questions are already of interest in our todays society and can also be related to normal human driven cars.

I can imagine much benefits of automated driving. Nevertheless it still has to be proven that autonomous driving is more reliable than human driven vehicles because there is no statistically significant data to back up that claim.

While it is important to keep the discussed question on how an autonomous vehicle should behave in a trolley problem like situation, the effort should be used to prevent the arise of such situations beforehand. Any other focus would not be morraly responsible as it accepts that there should be unavoidable harm in some situations, whereas it is our obligation to avoid this situation entirely as best as possible.

Bibliography

[A$^+$13] National Highway Traffic Safety Administration et al. Preliminary statement of policy concerning automated vehicles. *Washington, DC*, pages 1–14, 2013.

[AG17] Deutsche Bahn AG. Autonome elektrobusse. `http://www.deutschebahn.com/de/Digitalisierung/autonomes_fahren_neu/autonome_elektrobusse.html` as of 16.2.18, 2017.

[Alc05] Blake Alcott. Jevons' paradox. *Ecological economics*, 54(1):9–21, 2005.

[BSR15] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. Autonomous vehicles need experimental ethics: are we ready for utilitarian cars? *arXiv preprint arXiv:1510.03346*, 2015.

[BSR16] J.-F. Bonnefon, A. Shariff, and I. Rahwan. The social dilemma of autonomous vehicles. *Science*, 352(6293):1573–1576, jun 2016.

[Bun] Statistisches Bundesamt. Verkehrsunfälle zeitreihen. `https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/VerkehrsunfaelleZeitreihenPDF_5462403.pdf?__blob=publicationFile`.

[Fed16] Hannes Federrath. Eine telefonnummer ist ausreichend, um eine person mit einer drohnen-rakete zu treffen. `https://netzpolitik.org/2016/informatik-gutachten-eine-telefonnummer-ist-ausreichend-um-eine-person-m` as of 17.2.18, 2016.

[fVudI17] Bundesministerium für Verkehr und digitale Infrastruktur. Bericht der ethik-kommission automatisiertes und vernetztes fahren. `http://www.bmvi.de/SharedDocs/DE/Publikationen/G/bericht-der-ethik-kommission.pdf?__blob=publicationFile` as of 19.2.18, 2017.

[GM17] Jan Gogoll and Julian F Müller. Autonomous cars: in favor of a mandatory ethics setting. *Science and engineering ethics*, 23(3):681–700, 2017.

[Goo14a] Noah Goodall. Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board*, (2424):58–65, 2014.

[Goo14b] Noah J Goodall. Machine ethics and automated vehicles. In *Road vehicle automation*, pages 93–102. Springer, 2014.

[HNR15] Alexander Hevelke and Julian Nida-Rümelin. Responsibility for crashes of autonomous vehicles: an ethical analysis. *Science and engineering ethics*, 21(3):619–630, 2015.

[Jar17] Katharine Jarmul. Deep learning blindspots. `https://media.ccc.de/v/34c3-8860-deep_learning_blindspots` as of 11.3.18, 2017.

[Lin14] Patrick Lin. What if your autonomous car keeps routing you past krispy kreme. *The Atlantic*, 22, 2014.

[Lin16] Patrick Lin. Why ethics matters for autonomous cars. In *Autonomous Driving*, pages 69–85. Springer, 2016.

[MCM+17] J Manyika, M Chui, M Miremadi, J Bughin, K George, P Willmott, and M Dewhurst. A future that works: Automation, employment, and productivity. mckinsey global institute, 2017.