The morality of autonomous vehicles

Why forsaking manual driving can be considered a supererogatory act

Harmen Janssen

Masterthesis Ethics of Business and Organisations
28-02-2018
Course 799402
SNR 1279419
ANR 749248

Content

1. Introduction

As technology becomes more and more part of our lives we also increasingly depend on it to help secure our safety in our daily activities. Think of smoke detectors, microwaves and airbags and try to imagine our current world without them. However, in our modern cars the safety measures are no longer limited to airbags, automatic emergency lights and anti-lock braking system (ABS). Companies such as Google promise us a future in which cars are completely automated, no longer needing any human driver in any form, increasing road safety dramatically [1]. In this paper I am going to ask if the usage of autonomous vehicles leads to morally preferable situations and if so, if its usage can be considered either a moral obligation or a supererogatory act? Autonomous vehicles have the expected ability to improve the situation on the roads for us, but the question this paper will answer is to what extent these benefits are worth it. Of course, any discussion about the implication of automated transport is utterly irrelevant if the technology does not exist or is not generally available. Therefore I will first give a brief overview of the current development in the field, and the subsequent assumptions that are drawn for the sake of this paper.

In 1979, the Tsukuba Mechanical Engineering Lab was the first to build a vehicle that could drive up to fifty meters, in between two white lines painted on the road, independently and up to speeds of twenty miles per hour (Forrest & Konca, 2007). Since then a number of universities, companies and governments have gotten involved in the improvement of autonomous vehicles (which from here on will be referred to as AV or AVs for plural) leading to 38 years of developments, we now see examples such as the South Korean government which has just announced that they plan on building a 360.000 square meter test track for the development of AVs (Hawkins, 2017), Google has officially started offering rides in driverless minivans (Stewart, 2017) and Mercedes has claimed that they will have fully automated taxi's on the road within three years (Davies, 2017)

The continued and increasing involvement and development of AVs can be considered a testament to the industry's fate in the future of AVs. Research groups have published

[1] https://waymo.com/journey/

reports on the matter, claiming the majority of traffic may be fully automated as early as 2030 (Arbib & Seba, 2017). For the sake of this paper I will therefore work under the assumption that in the foreseeable future practical obstructions such as the commercial availability and the price of AVs for private individuals will be no greater then they currently are for 'regular' vehicles. That is to say, the argument that an AV might not be a viable option for someone because they may not be able to afford one will not be addressed in this paper.

The increased interest and discussions on AVs also means that there is an increase in many fields of study surrounding the topic. Within philosophy, particularly, a strong focus seems to be on the ethics of the choice the car has to make[2]. In this popular discussion two assumptions are essentially made, namely that even fully automated vehicles may still end up involved in a life or death accident, and secondly that the choice the vehicle makes in these situations needs to be pre-programmed in some form. It is precisely on this choice that discussions focus the most.

If we were to rephrase this question, at the risk of oversimplifying, we could ask ourselves; when do we let the car kill its user, or when do we let it kill those around us? One of the major academic institutions supporting this discussion is the Massachusetts Institute of Technology. On their website, moralmachine.mit.edu, they have created a platform to create better understanding and make the topic more accessible to the public. Or, as they put it themselves, "…providing a platform for 1) building a crowd-sourced picture of human opinion on how machines should make decisions when faced with moral dilemmas, and 2) crowd-sourcing assembly and discussion of potential scenarios of moral consequence"[3]. One of the published papers linked on the site discuss research into "Participants' approval of passenger sacrifice", "…moral preference for utilitarian AVs programmed to minimize the number of casualties" and questions such as "…how likely they would be to buy an AV programmed to prioritize protecting its

[2] Examples of which would be; http://dailynous.com/2017/10/06/philosophers-awarded-500000-study-autonomous-vehicles/, https://www.metro.us/news/the-big-stories/self-driving-car-ethics and https://www.timeshighereducation.com/opinion/philosophy-and-driverless-cars-kant-is-my-my-co-driver

[3] http://moralmachine.mit.edu/

passengers, even if it meant killing 10 or 20 pedestrians" (Bonnefon, Shariff, & Rahwan, 2016, pp. 2, 3)

Although I wholeheartedly agree that this very interesting line of questions is one that needs to be discussed in order to create a single unified platform for change, I would rather turn my focus to the topic of the 'regular' usage of AVs, not just the worst-case scenario. In this paper I will therefore discuss what it means to use an AV from a philosophical point of view, and furthermore show that for some users this transition comes at such a high cost that it can even be considered to be a supererogatory act. To achieve this, I will first summarize the assumed benefits of AVs, as they seem to be generally accepted. Afterwards I will discuss the benefits of manual driving, the moral aspects involved in both sides of the topic of AVs. The conclusion will show that the transition towards automated driving focuses too much on the expected benefits and fails to recognize a number of complications that drivers may experience during a transition to AVs.

2. Moral value of automated vehicles

To better explain what moral theories can be applied to the case of AVs, first an overview of the empirical research will be given. Afterwards two major philosophical theories will be introduced and explored, namely consequentialism and deontology. It is important to note that these two will not be fully analysed and explained for the same reason that this paper is limited to only two major schools of thought; the aim is not to provide the best matching normative theory but to demonstrate the concept. That is to say, it is an example of how these theories can be applied to a case they currently are not.

2.1 Empirical research

The aim of this paragraph is to set or show a standpoint, which can be considered to be generally accepted in our society, namely that AVs will have a positive impact by significantly decreasing casualties in traffic accidents and bringing with it several other benefits.

2.1.1 Level 4 automation

In 2016 the RAND Corporation published an in depth analysis of the changes that may be brought about by AVs. RAND is a research organization that develops solutions to public policy challenges that receives funding from governments, the private sector and private individuals, employing staff with more then 600 doctorates and almost 400 masters' degrees, arguably making them a suitable authority in the field. In the report, the authors adhere to five levels of autonomy as set by the National Highway Traffic Safety Administration (NHTSA), an American governmental organisation. I will first summarize the distinctions made as they are intensively used in their analysis. The NHTSA ranks automation as follows:

-Level 0, no automation;

- *"The driver is in complete and sole control of the primary vehicle functions (brake, steering, power and motive throttle) at all times, and is solely responsible for monitoring the roadway and for safe vehicle operation"*(Anderson et al., 2016, p. 2)

-Level 1, function-specific automation;

- *"Automation at this level involve one or more specific control functions; if multiple functions are automated, they operate independently of each other* (Anderson et

al., 2016, p. 2). However, at this level the driver still has overall control and responsibility. Specifically, either the driver can give, or the vehicle can take limited control when needed. Lastly the vehicle may aid or contribute to the control the driver has over the vehicle. An example of this would be cruise control.

-Level 2, combined-function automation;

- At this level two or more primary control functions are automated and designed to work in unison in order to take the function away from the driver. However, the driver remains responsible for monitoring the roadway and can be expected to take back control without prior warning.

-Level 3, limited self-driving automation;

- At this level the driver can hand over all functions needed for safe driving to the vehicle under certain traffic and weather conditions. The driver can rely heavily on the vehicle to monitor changes in those conditions but is expected to take control occasionally, albeit with a confortable transition period.

Level 4, full self-driving automation;

- At this point the vehicle is designed to preform all functions and responses. The 'driver' may input a destination but is not expected or even supposed to take control of the vehicle at any time during the trip. *"This includes both occupied and unoccupied vehicles. By design, safe operations rest solely on the automated vehicle system"* (Anderson et al., 2016, p. 3).

The RAND Corporation firstly uses the transitions between the levels to show past increases in road safety were attained by different forms of automation. They argue for instance that dynamic brake control, a level 1 feature, does not help prevent driver error. It only helps the driver once he has made his choice to break. What I aim to discus in this paper however, are instances in which the outcome of the situation on the road is no longer dependent on the driver's response. This form of automation then starts at level 3 and 4, the levels at which the driver can hand over full control to the vehicle.

The authors note that 14.000 out of the 32.000 deaths caused by car-accidents in America over the course of 2011 (Anderson et al., 2016) were cases of single-car accidents. Even without taking multiple-car accidents into account (because other cars

might not have had similar systems) it is easy to see the potential benefits of automation. When we consider the transition to level 4, and we assume that this level of automation *is* safer because control is taken away from the driver, we also tackle alcohol related accidents. These account for 39% of accidents with fatal outcome over the same year on American roads (Anderson et al., 2016). Level 4 automation could then arguably prevent a third of all deaths in traffic accidents.

2.1.2 Secondary benefits

RAND furthermore expects a number of indirect benefits in their report such as a reduction in pollution, a reduction in congestion and an increase in the usage of alternative fuels (Anderson et al., 2016, pp. 24, 28, 33, 36). Specifically, the report states that nearly twenty percent of all greenhouse gas emissions and around sixty percent of all petroleum use can be attributed to 'light-duty' passenger vehicles. To elaborate, three different areas in which AVs might impact energy and emission are specified, namely; fuel efficiency, carbon-intensity and life-cycle emission, and change in vehicle miles travelled.

Firstly, on fuel efficiency, there are expected benefits that are already being experienced by level one, two and three automation. Specifically, functions such as cruise control help vehicles 'run smoother'. That is to say, this so called 'eco-driving' tells drivers when to change gears and helps them accelerate and decollate more gradually. It is expected that further automation may increase vehicles ability to perform 'eco-driving' and may improve benefits in fuel economy by four to ten percent (Anderson, et al., 2016, p. 29). Moreover, uptimized driving through automation leads to better positioning on the road, which in turn allows for an increased travel lane capacity and reducing fuel wasted during congestion of roads.

Other improvements may be made in the design of the vehicle. A scenario is considered, for instance, in which the weight of cars can be reduced significantly because there are hardly ever any accidents. Lighter vehicles require less power and consume less fuel, but the report rightfully notes that "… the realization of these benefits will require AV consumers to have confidence that accidents with non-AVs are also avoided, which is likely to limit the types of substantial weight reduction to Level 3 or Level 4 automation

and will depend upon nearly universal adoption of this technology so the risk from non-AVs is minimal" (Anderson, et al., 2016, p. 31). This may be taken to mean that the ammount of benefits that can be expected from AVs is directly dependent on the percentage of vehicles on the road that are autonomous. Because these vehicles communicate with each other, more vehicles to communicate with will lead to better results. More on this will be discussed further on in the paper.

The benefits discussed above will in turn have consequences on the technology used to propel our vehicles. If AVs indeed help to reduce the weight of vehicles because there are fewer 'heavy' safety measures needed, we may consider for instance, having lighter vehicles mean electric cars require smaller and lighter batteries. Even further improving on this, if AVs help with vehicles drive cycle, driving more efficient and reducing the energy needed to reach our destination, batteries may be even smaller and lighter. RAND foresees that this may not only help in reducing emission, but also in reducing prices for consumers (as batteries are a very costly part of electric vehicles) but may also help speed up the transition to electric vehicles because of this. Furthermore, this line or reasoning leads us to conclude that the usage of smaller batteries may significantly decrease the life-cycle cost of producing them and reduce the environmental impact at the end of their lifespan (Anderson et al., 2016, p. 34).

The expected results of AVs as discussed above, are subsequently referred to as the *benefits* of automated driving. This report however, does not discuss the moral implications of these consequences, nor does it even confirm that these 'positive' results are morale preferable or best. In the next chapter we will do precisely that.

2.2 On the morality of automated vehicles

When discussing the morality of machines in the cases of AVs the topic of discussion quickly shifts towards the dilemma of cars choosing who lives and who dies. As explained in the introduction, this subject is currently discussed most when debating the morality of AVs and is easy to link to morality in general. An argument against the topic of this paper would be that in leaving out the 'choices' the car has to make in certain undesirable situations a large part of the moral or ethical discussion surrounding AVs is

ignored. I argue that this is not the case however, and instead show that the 'choice' that needs to be made shifts to a different act as the following examples show.

In his work (Moral Enhancement and Freedom, 2011) John Harris argues that humans need the ability to fail in order to be good. He argues that in absence of this freedom to fail there is no choice to be good, and therefore the virtue disappears as well. When we discuss the level 4 automation, the 'driver' of the vehicle no longer has the ability to impose any form influence on safety of the car on the road, and with it loses all possibility to fail. This influence arguably then shifts and comes to lie with the programmers of the car and those they take their orders from, respectively. I agree with Harris's premise that the driver, by lack of influence during a crash, can no longer do 'the right thing' in such a situation. However, I argue that the freedom to fail is still there, it simply shifts to 1) buying an AV or 2) using an AV when available. If the expected benefit of using an AV is an overall reduction of casualties in traffic, people who have to choice between either using an AV or a non-automated vehicle still have the ability to fail if they choose not to go with this option.

Secondly, an objection that could be raised against the subject of this paper is that this new technological development is one that is almost unavoidable or in some way, a choice that is going to be made for us. In the same paper (2011) Harris states that "One thing we can say with confidence is that ethical expertise is not 'being better at being good', rather it is being better at knowing the good and understanding what is likely to conduce to the good" (Harris, 2011, p. 104). When applied to AVs I believe Harris would say the automation does not entail any form of moral improvement, that is to say it does not make us morally better per se. Instead we should first recognize the possible benefits and secondly do not fail to act on this. This second point might be problematic according to Harris, as we know how "…lamentably bad we are at doing what we know we should" (Harris, 2011, p. 104). This does not fully stem from a lack of moral resolution but from a conflict in our many purposes and priorities. This, in my eyes, is a point on which the subject touches on morality. While priorities give order to what morally right thing you may or may not do (first), obligations imply (or rather impose) a 'jump' to the top of that list.

To investigate this 'other side' of the morality of AVs further, a number of ethical theories will be discussed in this chapter. Different forms of ethics generally support different views on what is right and wrong. In this chapter I will discuss what implications AVs may have on us when viewed from both consequentialist and deontological moral standpoints, or in other words, when one focuses on either the principles for actions or its consequences. It is important to note here that the aim is not to completely summarize the ethical debates or the entirety of the philosophical field of morality. Instead, the aim of this chapter is to provide the reader with a clear general image, along with a specific example of instances in which philosophical normative theory can be applied to the case of AVs.

2.2.1 Consequentialism

Firstly, consequentialism is a theory that focuses on the future to determine whether something is right of wrong. That is to say, it is teleological as the rightness or wrongness of an act is solely determined by the results of the act. The Stanford Encyclopedia of Philosophy describes its moral theory as one "…which holds that whether an act is morally right depends only on the consequences of that act or of something related to that act, such as the motive behind the act or a general rule requiring acts of the same kind." (Sinnott-Armstrong, 2015). An example of this can be the taking of one life to prevent the death of many others. Although many would argue taking a life is never the right thing to do, a consequentialist would argue that so long as the consequences of this act promote a greater good, it is the right thing to do. This, as a general concept, still leaves a number of questions to be answered. For instance, one may ask if our acts can be considered to be morally right if the outcome is not as intended. That is to say, what if the aim does not match the outcome and unintentional good is caused, or the sought after benefits are not achieved? Similarly, it is a theory that focuses on the greater good. However, this leaves the question of what the best consequences are when faced with the choice, open for debate and in need of a way of 'ranking' consequences. An example that springs to mind in the context of this paper for instance, would be whether people being entertained (and fulfilling their desires) or being safe would be the 'better' consequence of an action. Alternatively, the question can be raised whether me expressing my freedom may come at the risk of those around me.

To better determine which consequences can be considered to be the best, a further distinction is made within consequentialism. Specifically, the question of who it is this 'good' happens to. For the sake of this enquiry, this paper will not focus on ethical egoism or ethical altruism. Instead, the concept of utilitarianism will be further explained. Utilitarianism most easily describes 'good' as that which, as a sum of all consequences of action, produces a better outcome than any alternative action. John Stuart Mill defends one specific form of utilitarianism in his book *Utilitarianism* (1979). In his work, Mill describes the basic principle of utilitarianism as what he refers to as the Greatest Happiness Principle, or the principle that holds that "…actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness" (Mill, 1979, p. 14). To elaborate, Mill defines happiness as pleasure or the absence of pain. Alternatively, unhappiness is defined by pain and the deprivation of pleasure. Mill admits that this definition of utilitarianism is one that is not fully complete, as it does not tell us exactly what is meant with the terms pleasure and pain. However, he does stress the importance of knowing that, even without further specifying what can be considered pleasurable or painful, the core concept is knowing that the only ends we truly seek are pleasure and absence of pain. All other ends we pursue are sought after solely because of the pleasure they bring or the pain they allow us to avoid.

One of the benefits discussed in chapter 2.1 of this paper is the expected decrease in traffic accidents, and as a further consequence the decrease in casualties as a direct result of any such incidents. If we take the basic utilitarian principle as described above, I can argue both in favour and against the usage of AVs. One could argue, for instance, that the joy of driving a fast car is a pleasure worth pursuing as it brings us happiness. On the other hand, the possibility of being involved in an accident, injuring yourself or loved ones can be avoided to a certain degree by opting to use an AV. Both of these examples could be considered to be 'good' actions according to utilitarianism. However, a 'better' of the two needs to be found. Mill agrees and writes, "It is quite compatible with the principle of utility to recognise the fact, that some kinds of pleasure are more desirable and more valuable than others" (Mill, 1979, p. 16). Whereas the quantity of a pleasure lying simply in it being greater in amount and can therefor be more systematically weighed, the quality of a pleasure is more ambiguous in its definition.

Mill states “Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure” (Mill, 1979, p. 16). To elaborate, Mill argues for a situation in which someone that is fully familiar with both pleasures is necessary in order to truly measure what quality is ascribed to either pleasure. Mill further specifies that in these cases, the pleasure that is chosen can be considered as the higher quality pleasure even if we know it to be accompanied by greater discontent and when we would not resign this pleasure even for a higher quantity of pleasure which may lead from the alternative choice. Revisiting the examples given in the previous paragraph, between the joy of driving and prevention of the pain of being involved in an accident, Mill’s concept can be further applied. The repeated and continues joy of driving that is achieved when driving seems quantitatively greater. However, if one has experienced both this pleasure and the pain of being involved in an accident before it would arguably be easy to see how one would choose automated driving.

Mill further elaborates, stating, “…those who are equally acquainted with, and equally capable of appreciating and enjoying, both, do give a most marked preference to the manner of existence which employs their higher faculties” (Mill, 1979, p. 17). It is in these higher faculties that a distinction is made. Mill further explains that this difference between beings of higher and lower faculties is one that is inescapable and that a being of higher faculties can never truly hope to sink towards a lower existence. Mill illustrates this point by giving examples, stating “…no intelligent human being would consent to be a fool, no instructed person would be an ignoramus, no person of feeling and conscience would be selfish and base, even though they should be persuaded that the fool, the dunce, or the rascal is better satisfied with his lot than they are with theirs” (Mill, 1979, p. 17). In the case of AVs this argument can be used to further distinguish which outcomes are the ‘better’ consequences. Specifically, a being of higher faculties (as all humans are, according to Mill) is more complex to satisfy, or rather needs more in order to find pleasure and as such have the ability to experience ‘higher’ pleasures. Moreover, they are disposed to more severe suffering and in many different forms.

This means we are in a sense confronted with our ability to 'know better'. When we face the choice to, simply put, choose pleasure (manual driving) or safety (autonomous driving) we are forced to recognize our ability to find suffering in many different forms. However, we already face this choice in a number of different leisure activities. Imagine for example, the sport of skiing. In which case, one might weigh the pleasure expected to be had in the act, and compare it to the risk of being injured and pain accompanied with it. Similarly, a case can be made for instance, when swimming with sharks, skydiving or climbing mountains. What makes these cases different from AVs, and driving any vehicle manually for that matter, is that it is not just your own wellbeing to consider. When you choose to take your car out on the road you involve yourself in a number of complicated traffic situations. Any one of which could lead to the injury of strangers, albeit in other vehicles or on the road. In the case of manual driving we have come to accept these risks in society, under the guidance of rules and laws. However, as mentioned above, Mill argues that the end to an action is always pleasure or avoidance of the absence of pleasure. Moreover, we've established that humans as being of higher faculties have the ability experience pain or the absence of pleasure in different forms and severity. Guilt for instance, is one of these states as it is one in which pleasure is absent and is actively avoided by most people by driving 'safely', obeying speed limits and waiting at red lights.

Lastly, Mill responds to the objection that there are a number of cases in which beings of higher faculties ignore these options and elect to go with lower pleasures. Why else would one choose to smoke when it is quite obvious that a person's health would be the greater good? The answer Mill, argues lies in men's 'infirmity of character'. It is because of this weakness that, even in knowing their choice is less valuable man still chooses the lesser. Moreover, returning to examples of swimming with sharks, jumping from airplanes and driving fast cars, Mill stresses this infirmity of character not only shows when the choice is between two physical pleasures, but just as much in a choice between the physical and mental pleasures, "They pursue sensual indulgences to the injury of health, though perfectly aware that health is the greater good" (Mill, 1979, p. 19). For example, consider why a person would ever go skydiving if health is the greatest good, while for others there is no greater joy and the risk is hardly even considered. This will be further discussed in CHAPTER five of this paper.

2.2.2 Deontological theories

Whereas consequentialism focuses on the consequences of our actions, deontology does essentially the opposite. For instance, we may feel that there are certain duties such as taking care of our family, which are enforced by our society. Normative deontology theory tells us what actions we could, should or have to do in these situations. This, in contrast to consequentialism, does not focus solely on the good created by the action performed. Instead, deontological theories, unlike consequentialism, are not solely focussed on the future but look both forward and backwards. Whereas consequentialism in theory could allow for any act to be performed, even for instance the taking of a life, for as long as the sought after effect is 'good', deontology does not permit such a justification. Instead, some deontological theories hold certain principles or duties as absolute, in that no matter what amount of morally good outcomes are created as a result of that action, they would still be considered forbidden. The right choice is then not determined as a weighing of options, the quality or quantity of the benefits. Instead it is in its conformity with moral norms. Moreover, something cannot be said to be 'good' as long as it is not firstly 'right', we can even go as far as to say, no matter how good an action may be, it should not be undertaken unless it is also right.

One specific deontological theory that will be discussed in this chapter is that of William David Ross as discussed in his book *The Right and The Good* (1930). Ross remarks that there always seem to be three key elements which are intimately related in any moral theory, namely that of right, good and morally good. Subsequently, he starts his definition of this moral theory by asking what the meaning of 'right' is. In reflection on the utilitarianism as discussed in the previous paragraph, we might say that an act is always performed for our own pleasure, or alternatively the absence of pain. Ross does not necessarily disagree with pleasure as a 'good', but stresses that this is not as important in determining the rightness of the act (Ross, 1930, p. 16). For as soon as an act is done for ones own interests it is not done out of any concept of rightness but out of self-interest. Furthermore, Ross argues that it is not always in the future that the rightness of an act is determined. He gives the example of a man fulfilling a promise he had made earlier. This man does not fulfil his promise because he thinks he should do so, or because the total consequences of doing so outweigh *not* doing so. He fulfils the promise less so because of what might happen in the future, and more because of what

happened in the past. As Ross puts it, "What makes him think it right to act in a certain way is the fact that he has promised to do so —that and, usually, nothing more" (Ross, 1930, p. 17).

Ross proceeds of this bases of what is right and distinguishes between eight 'prima facie duties'. These duties, he remarks, should not be seen as duties in themselves, but as a means of referring to a characteristic an act has, for as long as it is not overridden by a 'proper duty', a strong presupposition in favour of preforming this act exists. That is to say, these eight duties are not considered to be proper duties, only that which can be considered to be a moral obligation would be a proper duty. Ross distinguishes the two in stating "…what I am speaking of is an objective fact involved in the nature of the situation, or more strictly in an element of its nature, though not, as duty proper does, arising from its whole nature" (Ross, 1930, p. 20). The concept of moral obligations will be further discussed in paragraph 4.2. Instead, first a brief explanation of Ross's prima facie duties will be given and applied to the usage of AVs.

The first two duties stem from actions we have taken in our past. Specifically, the duty of fidelity, which entails the keeping of a promise made, and the duty of reparation, which steers someone to make amends for wrong previously done unto others. The third duty also lies mostly in acts from the past, but this time in acts done by others to me. Ross calls these duties those of gratitude. In chapter 2.1 of this paper the possible decrease in emissions that accompany the usage of AVs was explained. I argue, for instance, that our nations promise to cut pollution or a willingness to *restore* our environment for future generations by choosing more environmentally friendly transport (and therefore possible AVs) is a choice driven by these first two duties as discussed by Ross.

The fourth duty lies in a discrepancy between wellbeing and worthiness of this wellbeing. Ross states, "Some rest on the fact or possibility of a distribution of pleasure or happiness (or of the means thereto) which is not in accordance with the merit of the persons concerned..." (Ross, 1930, p. 21). If such a person of merit, worth or value does not have (the ability to create) happiness such as he is deemed worthy of, one should feel a duty to correct this, a duty of justice. When looking at the case of AVs, general road safety affects everybody, not only the vehicles users but also those who share the public

roads. In this case it is clear that the best possible physical state or even more basically, the safety to walk the streets with minimal risk of being injured could be considered merited for anybody.

The fifth and sixth type of duties discussed are those of beneficence and are more directly applicable to the usage of AVs. Ross separates these duties into two kinds, namely, the duty of beneficence and the duty of non-maleficence. Firstly, it is our duty to improve, or do 'good', onto other beings, for instance in respect of virtue, intelligence or pleasure. These acts could be summarized as doing 'good' but are not fully comprehensive of the concept, Ross argues. Specifically, the example of injuring someone may constitute a failure to do 'good' to others. However, this negative connotation shows not so much a duty of beneficence but one of non-maleficence, which Ross defines differently. In his work he states "…it is really the duty to prevent ourselves from acting either from an inclination to harm others or from an inclination to seek our own pleasure, in doing which we should incidentally harm them" (Ross, 1930, p. 32). Ross uses the negation of the concept to 'not inflict harm' to show that this not only means that someone is not the *cause* of harm. Instead he argues that in having the ability to prevent harm to others even when that person is not the cause should be acted upon. This is also described as the duty of harm-prevention.

Moreover, when applying these duties to the case of AVs, they speak in favour of their usage. The duty of beneficence seems to lie mostly in our ability to reduce emissions and environmental impact. If switching to AVs may contribute in some way to returning our 'damaged' environment to its original state, this would be a beneficent act, not so much towards our fellow members of society but more so for the future of its future members. Alternatively, on the duty of non-maleficence, one could make a case for switching to AVs as to *not* do so would be choosing your own pleasure over the safety of others, in which case you have a duty to prevent yourself on any such inclinations.

The final duty is one aimed at the self, specifically that of self-improvement. In this case, Ross argues, we have the ability to improve our own condition, and subsequently gives us the examples of 'virtue' and 'intelligence' (Ross, 1930, p. 21). This duty of self-improvement could furthermore be taken to also apply to aspects such as happiness,

moral standing, health and safety. In the case of AVs there is a clear contradiction between some of these aspects of self-improvement. For instance, while AVs would directly promote your safety and moral standing as you are less likely to be injured or injure others, it may also be so that your happiness is severely diminished as there was a lot of joy for you to be had in driving.

3. The cost of automated driving

Having discussed how two major extensive and more specifically a number of more specific normative theories could be applied to the choice of switching to AVs, it would seem that it is almost always morally best to do so. There is however, a different side to this argument. In this chapter different view will be introduced on the benefits of manual driving and will subsequently be explained and analysed.

3.1 On manual driving

One of the phrases that are quite often used when discussing AVs is that vehicles will become 'smart'. This not only refers to the cars capability to 'think' about the situations it encounters on the road, but also its ability to communicate. AVs are expected to exchange information with other AVs on the road and traffic systems such as traffic lights and other object in its environment. This vehicle-to-vehicle communication helps further increase safety as well as traffic efficiency on the roads (Giarratana, 2016). This form of communication between vehicles may help to prevent accidents before they happen, but is most beneficially used if all vehicles on the road use this technology. That is to say, public roads with exclusively AVs will be safer than a public road with mixed traffic and as such, all manual vehicles should be prohibited.

3.1.1 Cars and emotion

Alternatively, a case can be made against abandoning our manual vehicles altogether. Mimi Sheller (Automotive Emotions, 2004) argues that our current form of decision-making in the automotive industry overlooks the cultural, social, material and affective dimensions of driving. Sheller states that our current automotive culture "… is implicated in a deep context of affective and embodied relations between people, machines and spaces of mobility and dwelling, in which emotions and the senses play a key part" (Sheller, 2004, p. 221). To elaborate, Sheller argues that vehicles impact our lives on a much deeper level than simply practical use. Furthermore, this deeper impact is not sufficiently recognized when discussing the development of AVs. It is a subject that discusses the future of car culture, and with it what Sheller calls the "coercive freedom of driving" (Sheller, 2004, p. 221) that shapes the public and private surroundings of its users. She uses the term coercive here to illustrate how large of a role cars have taken in our society and moreover how dependent we have become of

them. When discussing the previously mentioned potential benefits of AVs, we are often inclined to simply conclude it must be the better option. Sheller's arguments, on the contrary, can be used to argue that this form of decision making is erroneous, as "… most practical efforts at promoting more 'ethical' forms of car consumption have been debated and implemented as if the intense feelings, passions and embodied experiences associated with automobility were not relevant" (Sheller, 2004, p. 222). To be clear, Sheller states that in discussing how to be more ethical in our usage of vehicles and prompting users to go for these options, we often neglect to discuss the way these choices impact most other aspects of our live. For example, over the last five years the Dutch government promoted the use of plug-in hybrid vehicles (which have to be charged after use) by offering exemption from road tax and registration fees for companies that lease hybrid vehicles. However, as a large part of the drivers did not actually charge their electric batteries and continued to use their regular petrol engines the road taxes for these vehicles were quickly raised from zero to four and seven percent[4].

Sheller takes her argument to mean that car consumption is not a case of rational economic decision-making. Instead she includes the emotional and sensory responses to driving, the patterns of work and social life, kinship and habitation. The usage of a car then has a deeper, more intimate relationship in our culture and social activities, and as Sheller defines, "… an emotional agent is a relational entity that instantiates particular aesthetic orientations and kinaesthetic dispositions towards driving" (Sheller, 2004, p. 222) which leads her to conclude that "Movement and being moved together produce the feelings of being in the car, for the car and with the car" (Sheller, 2004, p. 222). It is in the latter, the aspect of being 'with the car', that an argument can be made against the use AVs. Sheller proceeds to explore four dominant sections of car consumption, namely that of 1) feeling the car, 2) being (in) the car, 3) family cars, caring and kinship and 4) national feelings about cars. In the next paragraph s I will briefly summarize the definitions given, and determine whether or not its implications speak in favour or against the usage of AVs.

[4]https://www.rvo.nl/sites/default/files/2016/05/5%20Years%20of%20Hybrid%20and%20Electric%20Vehicles%20living%20lab%20projects.pdf

The first argument Sheller discusses is introduced with the following citation; “Whilst I am driving, I am nearly always happy. Driving towards virtually anywhere makes me excited, expectant: full of hope.” (Pearce, 2000. p163) Drivers do not all experience the feelings cars equally. Some choose a vehicle purely for the sake of its function; others spend their fortunes on the most lavish and extravagant vehicles. However, all drivers develop some form of feeling for the car, its function and the relation between the two. As mentioned before, it is not solely the feeling of being in the car or for the car; rather there is also an element of being with the car. When we discuss the implementation of level 4 automation, most, if not all interaction with the vehicles primary functions is removed. The ‘driver’ in that case arguably is no more ‘with the car’ than you or I would be ‘with the bus’ when using public transport.

A second consequence of the ‘disappearance’ of this dimension may be strongly felt in consumerism and distinctions between brands. Sheller writers on the experience of seeing a car in a showroom;

> "Touching the metal bodywork, fingering the upholstery, caressing its curves, and miming driving ‘with all the body’ suggests the conjoining of human and mechanic bodies. Of course, viewing cars as prosthetic extensions of drivers’ bodies (…) is the standard fare not only of motor shows and advertising, but also of youth cultures, pin-up calendars, pop lyrics and hip-hop videos.” (Sheller, 2004, p. 225).

In the examples Sheller uses here it becomes clear once more that the car in general has achieved a role in the establishment of culture in which confidence, personality and even aspects such as sexual desirability are directly attributed to person based on the vehicle they own and operate. Furthermore, a drastic shift in the role, or even the emotional value of vehicles could then impact the way our social network interacts with us and in what light they view us. More on this will be discussed in the next paragraph.

On the second dimension, being (in) the car, once again Sheller starts her argument with a quote that illustrates the dimension she is going to describe at an almost emotional level to the reader; “It felt alive beneath my hands, some metal creature bred for wind

and speed. (…) It ran like the wind. I ran like the wind. It was as though I became the car, or the car became me, and which was which didn't matter anymore" (Mosey, 2000, p. 186). Admittedly, the setting might not be instantly recognisable to all at first, however it is not hard to imagine the feeling of being 'connected' to an inanimate object in a similar fashion. Sheller elaborates and argues for the existence of reciprocity between the environment and the organism. She asks in what way the presence of the car impresses on us as their users. The way moving feels, the way the steering wheel feels in your hands, the smell of fuel and the noise the suspension makes when you hit a bump are all important. These sensations, movements and experiences create emotions, either positive or negative ones, but are never created just by virtue of a person being in a moving object. Instead it occurs because there is a relationship between the different dimensions, or as Sheller states, they "… occur as a circulation of affects between (different) persons, (different) cars, and historically situated car cultures and geographies of automobility" (Sheller, 2004, p. 227). This then, is a complete package. Sets of variables that help construct a particular sort of feeling, different for each individual. An experience that speaks to the emotional relations people have.

In the third and fourth dimension Sheller discusses family, caring, kinship and the national feelings and identities that exist about cars. In discussing this, Sheller touches on the subject of anti-car protest, specifically on the duality in moral arguments used by those who oppose the usage of (certain) cars. In these cases, there is "… a conflict between an ethics which is concerned with aggregate effects of personal action on the world at large and a morality that sees caring in terms of more immediate concerns such as one's partner and children" (Sheller, 2004, p. 229). In other words, a person may strongly oppose to construction of new highways as to do so would damage nature and allow for further air pollution, but still be an avid car user on a daily basis. The usage of a car being a prerequisite in order to maintain their jobs, bring their children to school and maintain a social life involving many friends and family. Driving then, once again, is shown to be a set of experiences larger than just the functionality of transport. (More on this duality in morality will be discussed in the next chapter). On the national culture of cars, Sheller argues no two countries are the same.

> "Stereotypical 'Western' perceptions of driving in 'Third World' countries (…) rest partly on a clash between these different national styles, motorscapes and affordances. What image of Cuba is complete without the fading glory of the massive tail-finned cars from the heyday of US imperialism, lumbering zombies from a pre-revolutionary capitalist era?" (Sheller, 2004, p. 234).

I would argue that this statement shows us the complexity of the transition into an all-round use of AVs. As mentioned before, the 'intelligence' of the AV allows the car to communicate with other 'smart cars' on the road. In doing so, a large part of accidents can be prevented before they even happen. What follows, is that the greatest possible benefit in AVs is when there are solely intelligent, communicating cars on the roads. For a nation to achieve this does not seem entirely impossible, albeit a long and difficult transition. However, when we take into account that, for instance, not all of our neighbouring countries have the same economical and technological foundation or 'starting point' for change. Likewise, an argument can be made that cars and driving are more deeply rooted into the German culture than in the Dutch culture. This difference could largely influence willingness and international cooperation but may even feel as an attack on national culture. Imagine, for instance, The Netherlands public roads being used by exclusively autonomous vehicles. Would this mean non-automated vehicles visiting from Germany are no longer allowed? For a country such as the Netherlands this could never be realistically considered as international trade provides for much of the countries welfare. This question however, falls outside of the scope of this paper but will be revisited briefly in the concluding remarks.

3.1.2. Analysis

Sheller's work is, although directly relating to the subject of automotive mobility, not specifically written on the subject of automated driving. It is also written more from a sociological standpoint than a philosophical one. Some elaboration and analysis of her arguments is therefore done, based on the arguments that were provided in the previous paragraph.

Firstly, on Sheller's first dimension, namely that various drivers experience driving cars differently, and that emotions are therefore also experienced differently. Feeling the car

is a complete package. It comes with pleasure, the thrill of driving and the joy of the open road. However, it also comes with fear, frustration and pain. Sheller states, "The stomach-turning feeling of witnessing a car crash or the terrors and permanent anxiety produced by being in an accident are the dark underside of 'auto-freedom'" (Sheller, 2004, p. 224). The emotions that make you love your car and the joy of driving also give you the frustration you feel at other drivers on the road and anger towards the governments choice to have lowered speed limits. However, as these are two sides of the same coin, and AVs are expected to have a large, negative impact on the affectionate feelings people have towards a car, accompanied by a lowered sense of involvement with the vehicle and less joy in driving, it would also stand to reason that these frustrations, fears and anxieties would be reduced as well.

Secondly, the next dimension that was discussed was that of being 'in' the car 'with' the car. A question can be raised to what degree technological enhancements in cars serves to separate the driver and the vehicle, or whether they instead may help integrate the two further. If a part of the emotional experience of driving a car is the feeling of having a smooth gear transmission, a quick throttle response and an ergonomically pleasing steering wheel, how can the driver ever feel 'one with the car' if he no longer has these things? Admittedly, these experiences can never be exactly the same. However, as Sheller also points out, the experience is always shaped by a large number of circumstances albeit emotionally, economically, time and culturally related. The AV could then be considered an instigator of change in all these dimensions, and in some way creates a new form of 'connectedness' with a vehicle.

In her work Sheller talks about the shift in our embodied feelings of cars as using the vehicle as machines, to (partially) computerized control of the machine. In her writing, Sheller does not touch upon the subject of AVs directly. Instead she mentions the other safety features that were widely and quickly accepted by car users everywhere, such as seatbelts, airbags and crumple zones.

> "Features such as automatic gearboxes, cruise control, voice-activated entry and ignition, GPS- navigation, digital music systems and hands-free mobile phones all

> ‘free’ drivers from direct manipulation of the machinery, while embedding them more deeply in its sociality…” (Sheller, 2004, p. 229).

Once again, regarding ‘driving’ as a set of complex social, emotional, economical factors, one can see how the automation of driving may serve to further ‘link’ the user to his vehicle, or at least the total experience of using a vehicle. Whereas one dimension of the total experience may be reduced or completely taken away, another may be strengthened or added to this package of experiences. Sheller even acknowledges this herself, stating “Collective cultural shifts in the sensory experience of the car hint at what might be necessary were there to be a wholesale shift toward a new (more ethical) culture of automobility: a new automobile aesthetics and a new kinaesthetics of mobility” (Sheller, 2004, p. 229). This will be further discussed in the final chapter.

The final dimension that was discussed in the previous paragraph was that of kinship and national feelings towards cars. I argue that this dimension also raises the question as to who sets the terms in the transition towards AVs. Sheller touches on this subject as well in her work, stating “In considering these practices of national branding I do not mean to suggest that cultures of automobility will change simply by designing cars in new ways. Nor do I believe that it would be possible for a single nation (or multinational corporation) to lead the way in creating a more ethical car culture” (Sheller, 2004, p. 235). In this, we acknowledge a small, incremental development and an increasing willingness to experiment but are also forced to accept that no drastic transformation of vehicles and infrastructure have been made or seem to be happening in the near future. Likewise, car culture and car emotions only seem to move slowly in the direction of AVs, with no large opportunities up for the taking.

If we were to draw a conclusion based on Sheller work as discussed above, I would like to point out two things become apparent. Frist and foremost the fact that there is a deep and very much underappreciated emotional dimension to the usage of cars. This dimension is hardly every discussed when talking about the implications of AVs and in doing so, does not recognize enough what will be ‘taken away’ from drivers when they switch to AVs. Likewise, there is not enough recognition to the social, economical and cultural dimension that will be affected when a switch to AVs is made. As Sheller states it

> "Debates about the future of the car and road system will remain superficial – and policies ineffective – insofar as they ignore this 'deep' social, material and above all affective embodied context. Social research on automobility will also remain cramped in the 'transport studies' enclave until we recognize the full power of automotive emotions that shape our bodies, homes and nations" (Sheller, 2004, p. 237).

I believe then, that Sheller would not argue against any such shift towards AVs. She would however be right in pointing out the immense complexity well beyond the range of current discussion topic and academic focus.

4. Weighing the reasons

The aim of this paper is to show that transitioning to AVs, although morally praiseworthy cannot be considered a moral obligation and for some should even be considered to be a supererogatory act. To achieve this, the concept of moral obligation and supererogation will first be introduced, after which an analysis will be given.

4.1 Obligation and supererogation

So far this paper has introduced a number of expected benefits of large-scale usage of AVs, which were shown to have both (moral) advantages and disadvantages. However, as mentioned in the third CHAPTER when discussing Mill's vision on 'infirmity of character' of men, we do not always choose to do the required thing. Even though an option can be unmistakeably the best thing to do, there may be many reasons why we choose not to. Similarly, sometimes what seems to be the required thing to do may come at such a cost that we could not reasonably ask someone to do it. The aim of this chapter is to further discus what we consider to be moral obligations and what we regard as supererogatory acts.

4.1.1 Duty and moral obligations

In Ross's work, (as discussed in paragraph 2.2.2) he distinguishes between prima facie duties, and what he calls 'duty proper' or actual moral obligations (Ross, 1930, p. 20). To further discuss, a separate explanation will be giving on what the term moral obligation is taken to mean, for this we borrow from the definition of Stephen Darwall (Moral Obligation and Accountability, 2007). This theory should not be seen so much as a normative theory but as an understanding of the nature of moral obligation, which needs to be understood to further discuss on this topic. That is to say, I am not arguing that this specific theory is the only proper view on the subject. Instead Darwall's view on moral obligation is used because is serves as a good introduction to the subject. Darwall argues that morality can be seen as a working of equal accountability between rational beings in a society. To be specific, Darwall believes claims can be made of people but will only be accepted to the degree that other people have the authority to make claims on him/her. The acknowledgement furthermore lies in certain assumptions to which the people are committed simply by being a rational being, a part of a community.

The fact that, according to Darwall, obligation is then related to accountability also means that there is a consequence for not preforming an obligation. In our community we call something wrong only when we feel that the actor of that action ought to be punished. This may be through the law, through the opinion of others in this community or by ones own conscience (Darwall, 2007, p. 91). Furthermore Darwall believes that "There can be no such thing as moral obligation and wrongdoing without the normative standing to demand and hold agents accountable for compliance" (Darwall, 2007, p. 99). Your moral obligation comes from the fact that someone, according to other members of your moral community, can hold you accountable.

However, Darwall moves beyond simply the idea that being subject to moral obligations also means being subject to accountability from those with the normative standard to demand compliance. As mentioned above, Darwall argues that a system of mutual-accountability exists. He notes "... moral subjects must be assumed to be capable of imposing moral demands on themselves through recognizing that they validly apply to them as rational persons" (Darwall, 2007, p. 101). This leads Darwall to conclude that moral norms "... regulate a community of equal, mutually accountable, free and rational agents as such, and moral obligations are the demands such agents have standing to address to one another and with which they are mutually accountable for complying" (Darwall, 2007, p. 101). In other words, moral obligation is the demands we can reasonably hold others such as ourselves accountable for (not) doing.

Having introduced this concept of mutual accountability and general deontological theory, we now turn to a more specific and normative theory on obligations. Specifically, in this chapter we will discuss a number of aspects from Scanlon's contractualism as described in his work *What We Owe Each Other* (1998). If the basis of a moral obligation is indeed grounded in compliance, and whether it is justifiable for someone to (not) do something, then a theory of moral reasoning that leads to a general agreement needs to be discussed. Or to rephrase, if we are to hold each other equally accountable then there must be a shared ground to base this on. This is precisely what Scanlon does in his work. When discussing obligations we are essentially speaking of acts that are authorized or prohibited. This authorization can also apply to an entire class of action, and has an influence on the morality of the action. Specifically, isolated distinct instances of

preforming an act can have entirely different consequences than a common performance of similar actions. Scanlon argues;

> "As agents, if we know that we must stand ready to perform actions of a certain kind should they be required, or that we cannot count on being able to perform acts of another kind should we want to, because they are forbidden, these things have important effects on our planning and on the organization of our lives whether or not any occasions of the relevant sort ever actually present themselves" (Scanlon, 1998, p. 203).

To elaborate, consider for example if you were obligated to provide shelter for strangers during a storm. Knowing we hold this moral obligation, it would mean having to consider this possibility even when we are not specifically asked to preform this duty. For instance, it could mean buying a bigger house, an extra bed or more groceries. Similarly, the same can be said for those who experience the consequences of our actions. Scanlon would argue for instance that we do not experience privacy simply because my mail has not been opened when it is delivered to my house. Instead we experience the feeling of privacy because the mail has not been opened and that this action is considered prohibited, accepted as a general principle. Scanlon then argues that it is in this general acceptance of principles that we are all affected, summarizing; "…general prohibitions and permissions have effects on the liberty, broadly construed, of both agents and those affected by their actions" (Scanlon, 1998, p. 204).

Furthermore, acknowledgment of principles has additional consequences. "Because principles constrain the reasons we may, or must, take into account, they can affect our relations with others and our view of ourselves in both positive and negative ways." (Scanlon, 1998, p. 204). To elaborate, an example of a positive influence of principles will be considered. The principle of not violating ones privacy serves them by allowing them to be 'unobserved' when they want to, as an individual and helps them define themselves as independent persons. As such, they may choose to enter or avoid entering new relations as equals. If the principles we generally accept do not validate these reasons, our social interaction would be vastly different. Furthermore, Scanlon would even argue that without this validation of reasons it would even change the way we view

ourselves, possibly slowing our personal development or self-confidence. Consider for instance, if you would feel more or less inclined to interact with strangers who already know intimate details about your life, or if you would be more or less outspoken as a person if your words reached others than the intended audience.

This then leads us to the acceptance or rejection of these principles. This, Scanlon remarks, is not easily done, as "…an assessment of the rejectability of a principle must take into account the consequences of its acceptance in general, not merely in a particular case that we may be concerned with." (Scanlon, 1998, p. 204). As mentioned above, an action in itself, as an individual case may have entirely different principles then a general class of actions. In arguing that we should look at the consequences of general acceptance of a principle, Scanlon admits, we will be unable to do so as it is impossible to know which individuals will be affected and in what role they will have in the action. Therefore we cannot judge a principle on any individuals ends, personality or actions. "We must rely instead on commonly available information about what people have reason to want." (Scanlon, 1998, p. 204). Scanlon refers to this information as information about generic reasons and provides us with an example that suitably matches the topic of AVs as discussed in this paper.

> "We commonly take it that people have strong reasons to want to avoid bodily injury, to be able to rely on assurances they are given, and to have control over what happens to their own bodies. We therefore think it reasonable to reject principles that would leave other agents free to act against these important interests" (Scanlon, 1998, p. 204).

It is in this example that two conflicting reasons seem to clash when discussing the use of AVs, namely 1) avoiding bodily injury and 2) having control over what happens to their own bodies. In the first case, much in compliancy with Ross's prima facie duties (as discussed in paragraph 2.2.2), avoiding bodily injury to yourself and those around you serves as the strongest of arguments in favour of using AVs. When an agent chooses to use a manual vehicle, for whatever reason, that person then assumes a greater role in the lives of his fellow road users, as there is an increased chance of bodily injury. Alternatively, the second generic reason provided may be used to argue against the use

of AVs. Having control over what happens to you can be taken as an argument against having any choices forced upon you. If the principle of privacy entails the reasons that allow us to be equal individuals then the principle of freedom can arguably entail the same reasons. It allows us to enter into new relations as we choose them, not as we are forced to accept them as a result of immobility, which in turn lets us determine who we are as individuals. That is to say, we are no longer limited to working, studying and socialising within our immediate environment. Instead facing these choices on placing ourselves in different social settings and career paths, precisely because we have to freedom to do so, helps determine who we are as a person. This as well is very much in compliance with the arguments made by Sheller as previously discussed (in paragraph 3.1), stating driving cars is more than a functional thing and is part of a package of experiences that help us determine who we are.

These dimensions as described by Sheller and the generic reasons as described by Scanlon are not necessarily universally accepted. Not all agents are affected by the same principles in the same way and generic reasons are not simply defined as such because the majority of people share them. Moreover, Scanlon argues "If even a small number of people would be adversely affected by a general permission for agents to act a certain way, then this gives rise to a potential reason for rejecting that principle." (Scanlon, 1998, p. 205). This shows that, similarly to the theories of deontology and consequentialism, which were discusses in paragraph 2.2, there is both a need to look forward as well as backward in determining the best possible action to take. It can also be taken to conclude that transitioning towards AVs cannot be considered a moral obligation because a small number of people may be adversely affected.

Lastly, Scanlon argues that the ground for reasonably rejecting any principle ultimately lies in the costs for others and subsequently in what alternatives there are. To further discuss what reason people may have to inversely rank reasons and costs that we turn to Portmore's work in *Are Moral Reasons Morally Overriding?* (2008).

4.1.2. Supererogation

In addition to actions that can be seen as obligations, there are also actions that go beyond what can be expected. Or at least, that is what James Urmson argues in his work *Saints And Heroes* (1958). In it, he asks if an action can be morally good but not morally obligated. Once more it is important to remark that the work of Urmson, like other writers cited in this paper, was not chosen because his work is taken as a representation of the entire philosophical discussion on supererogation, nor because his theory is most accurate. Instead it is used because it serves as a good introduction to the subject and shows how philosophy can be used differently in the discussion on AVs. Urmson argues that in philosophical discussion of ethical theories there are generally three types of action recognized, whether it be explicitly mentioned or not. These three types are 1) those acts which are our duty, obligation or that which we ought to do, 2) acts that are right, that is to say, acts that are permissible from moral standpoints but are not required, and 3) acts that are wrong, that which we ought not to do. Urmson would argue however, that these three classes are not sufficient to include all types of actions. Specifically, Urmson argues, there is a type of action that is morally praiseworthy, but is not a duty and more than just a morally permissible act.

The type of action Urmson refers to is the type of action we may call heroic. Urmson further explains there are three scenarios in which we may call some a saint or a hero, out of which only one cannot be explained by the three classifications as mentioned above. Specifically, Urmson states;

> "… we may also call a person a saint if he does actions that are far beyond the limits of his duty, whether by control of contrary inclinations and interest without effort (…) we may call a person a hero if he does actions that are far beyond the bounds of his duty, whether by control of natural fear or without effort" (Urmson, 1958, p. 62).

To prove this type of action exists and cannot be placed in one of the three existing classes, Urmson gives us the example of a soldier that makes a split second decision and decides to throw himself on a grenade that threatens his comrades, and in doing so saves the lives of his fellow soldiers at the cost of his own. Of this action, Urmson argues,

we can clearly see that it was not the soldiers' duty to do so; it was something more than that.

> "Though clearly he is superior in some way to his comrades, can we possibly say that they failed in their duty by not trying to be the one who sacrificed himself? If he had not done so, could anyone have said to him 'You ought to have thrown yourself on that grenade'? Could a superior have decently ordered him to do it? The answer to all these questions is plainly negative" (Urmson, 1958, p. 63)

Urmson uses these questions as examples to show that the act itself does not fall within any of the three categories, a link to the previously discussed works of Scanlon and Darwall on moral obligation can be seen. Specifically, the negative answer to these question shows that no one could reasonably hold the soldier responsible for not performing the act, and the soldier himself (had he had the time) would have had sufficient reason to reject the choice itself.

Urmson considers the argument that the soldier may have seen it as his duty to throw himself onto the grenade. Urmson does not deny that this might be the case. Instead, he would argue that it might be just so that the action simply seemed like an obligation to the soldier, but that this does not make the act any less heroic (Urmson, 1958, p. 63). If the soldier had survived the act, he would still not have been able to tell anyone else it was his or her duty to do the same, or be told by anyone else that it was simply his duty to have done what he has done. Even in hindsight the heroic nature of the act would have been clear.

Furthermore, as shown in the citations in the pervious paragraph, Urmson then touches upon the subject of the sacrifices involved with choosing in what way the soldier acts. In the previous chapter a case has been made in favour of using manually driven vehicles, or in other words the sacrifices that one would have to make when switching to AVs. That is to say, in choosing the manual option in favour of pleasure, some safety is forfeited. Similarly, choosing autonomous driving in virtue of comfort and safety may arguably lead to a reduction of pleasure, which was previously gained, from driving. The problem with drawing a comparison between the example of the soldier and the

example of AVs lies precisely in these consequences to the action, and therefore needs to be further discussed.

Some people may regard driving as one of the greater pleasures in life, while others do not share this feeling. This difference between the people gives way to a discussion one would not as easily enter in the example of the soldier. The question can be raised whether there is really a noteworthy sacrifice to be made in choosing an AV specifically by those who do not share a passion for driving in general. Alternatively, the theories as discussed by Sheller would have many people argue that there is a substantial sacrifice. Moreover, Sheller's' theory shows that those who do not recognize that there is a sacrifice to be made in switching to AVs do not fully recognize the role cars have in our lives emotionally, culturally and socially. To be clear, I am not trying to compare the morality of the choice the soldier had to make to the choice that is being discussed in this paper. I am however, showing that for some people the type of action Urmson describes, even if it not objectively recognized by them as such, can be regarded the same.

In the introduction of this paper, an assumption is introduced and explained, stating that in the foreseeable future, accessibility and financial feasibility will be no bigger of an issue in the case of AVs than they currently are for manual cars. If this assumption holds true, the main argument against using an AV, which is largely covered by Sheller, might very well only apply a very small portion of drivers. As such, the question of whether or not we could reasonably ask someone to 'make the sacrifice' of switching cannot be universally answered as well as we may in the example of the soldier. Fortunately, universal agreement is not required in Urmson's theory in order for moral philosophers to make a distinction between different types of acts. Specifically, a majority of people would not hesitate and switch to AVs based solely on the expected benefits they were informed of. Subsequently, these people may (rightfully) consider themselves in the appropriate position to ask of others to do the same thing.

Although this paper has established that it would be morally permissible or even praiseworthy to choose an AV over a manual vehicle (chapter 2.2). Furthermore, a case has been made to show that this switch would bring with it a sacrifice for many people

and may influence drivers even in their social life and emotional and personal development. However, it would nonetheless seem that the act of switching from manual to automated driving could be considered supererogatory. This type of conflict of interests can already be found in our society in a number of cases. For instance, on smoking and gun laws discussions have been going on for years. In these cases the government has a large impact on what options there are and which seem best to choose. I would like to stress that this political question is not what is being analysed here, instead it is the moral question and the personal decision to act or not. What this approach of analysing the choice does show us however, is a sort of conflict, most basically, between an agents own interests and what morality asks of him. Portmore discusses precisely this subject: an agent's reason to promote his or her own self-interest.

In his work *Are Moral Reasons Morally Overriding* (2008), Portmore responds to those moral philosophers that recognize the existence of supererogatory acts and agent-centred options, stating that they are forced to recognize that "…the reason an agent has to promote her own interests is a nonmoral reason and that this nonmoral reason can prevent the moral reason she has to sacrifice those interests for the sake of doing more to promote the interests of others from generating a moral requirement to do so." (Portmore, 2008, p. 369). To elaborate, Portmore aims to investigate if the moral status of an act (albeit wrong, right or even heroic) is solely determined by the moral reasons for the act, or if the non-moral reasons may also attribute to the acts moral status. In determining whether or not the usage of AVs can be considered a morally right thing, I believe it is imperative this final theory on determining morality needs to be discussed.

The concept of agent-centred options refers the choice an agent faces as a being with moral status. Moral status entails that the agents interests have a certain worth and attribute something in determining what is morally required to do. Moreover, it also means the agent itself is an end, not just a means to bring about the morally best situation. Subsequently, this places the agent in the 'centre of the choice', that is to say, the agent is not obligated to always choose the greater gain over his or her own interest, nor are they obligated to always serve their own best interest. Portmore summarizes this in stating that this is "…the moral option of either promoting their own interests or

sacrificing those interests for the sake of doing more to promote the interests of others…" (Portmore, 2008, p. 369) and that in these cases we can recognize an act that goes beyond the call of duty in that "…in such instances, doing more to promote the interests of others is supererogatory" (Portmore, 2008, p. 369). In response to those who accept both agent-centred options and supererogation, Portmore aims to show that the role of nonmoral reasons is bigger then generally accepted.

In discussing whether or not moral reasons are morally overriding, Portmore introduces the concept of moral requiring strength and moral justifying strength. He explains moral requiring strength to mean that certain reasons that are normally accepted to *not* be taken, instead being morally impermissible to refrain from. Alternatively, moral justifying strength applies to acts that are morally permissible to perform that would normally be morally impermissible. That is to say, there are acts that one would morally not be expected to do, or even expected *not* to do, which may actually be morally permissible or impermissible depending on the moral requiring and justifying strength of the reason. Portmore uses these criteria almost like weights on a scale, determining whether one of two reasons has a higher moral requiring or justifying strength than the other. He states, a reason has more moral requiring strength if and only if the reason were to make it morally impermissible to do *anything* that the alternative reason would make morally impermissible to do, or if the reason makes it morally impermissible to do *some things* that the alternative reason would not make it morally impermissible to do (Portmore, 2008, p. 373). Alternatively, a reason has more moral justifying strength if that reason makes is morally permissible to do *anything* that the alternative reason makes it morally permissible to do, or if the reason would make it morally permissible to do *some things* that the alternative reason would not make it morally permissible to do. Portmore illustrates with the following example:

> "Even though it would be morally permissible to let an innocent person die in order to save one's daughter (as where both are drowning and one has only enough time to save one of the two), it would not be morally permissible to kill an innocent person in order to save one's daughter (as where one's daughter needs that person's heart to live). (Portmore, 2008, p. 373)

The example illustrates that, even though saving an innocent person from dying is generally considered morally required, a persons' reason to refrain from taking the life of an innocent person has far greater moral requiring strength than that of saving the life of an innocent person. Alternatively, Portmore gives an example of a woman faced with transferring her life savings to either purchase a new home or support a charity. In this example, Portmore argues, it would seem that all moral reasons would have her give her money away to charity. Furthermore, Portmore argues, that if the woman were able to do both, buy the house and help charity, the woman would be morally required to do precisely that. The difference between these two cases is then the cost to the woman personally, and while in both cases morality would seem to require her to transfer the money to charity, in the first case the nonmoral reason for not helping charity at her own costs would make it so that it would be morally permissible for the woman to buy the house. Here the reason has sufficient moral justifying strength, allowing for a nonmoral reason to prevent a moral requirement to be formed, even when the alternative has a large moral requiring strength.

Having explained his vision on reason and moral requirements, Portmore applies this theory to the concept of supererogation. Portmore generally agrees with Urmson's definition of supererogation as discussed in the previous paragraph. However, Portmore adds a criterion and stresses its importance, namely that a person only performs a supererogatory act if that person has more moral reason to perform that act over the alternative. In proving this claim, Portmore states that 1) someone who thinks they are doing the right thing (but is actually not) fails to properly appreciate the force of moral reason, and 2) that even though this person thought they were doing the right thing, the act they performed is not morally praiseworthy. Moreover, this leads us to conclude that 3) in order for an act to be morally praiseworthy, the agent must properly appreciate the relevant moral forces and 4) for an act to be more morally praiseworthy than another, there must be more moral reason to perform it. This then, in turn, is taken to mean an agent needs to have more moral reason to perform an act than any available alternative in order for that act to be considered to be supererogatory.

If moral reasons were to be morally overriding then by the definition discussed in this paragraph, supererogatory acts will always be morally required unless the reason lacks

moral requiring strength to make it a moral requirement. Alternatively, if moral reasons are not morally overriding we can imagine a situation in which the morally undefeated reason does not generate a moral requirement for performing the supererogatory acts because there is a nonmoral reason to perform an alternative act. Furthermore, having established that supererogatory acts go beyond what can be reasonably expected, and considering Portmore's statement that "…what explains the fact that it is morally permissible for the agent to fail to perform the supererogatory alternative is the fact that she has a sufficiently weighty nonmoral reason to perform some non-supererogatory alternative" (Portmore, 2008, p. 380), I argue that in assessing the morality of the choice either for or against the usage of AVs needs an account of both moral and nonmoral reasons. Nonmoral reasons here are reasons, which are neither considered to be moral nor immoral. One of the examples Portmore uses in his work an agent's reason for promoting her own self-interest (Portmore, 2008, p. 2)

Here a link to the transition to AVs can be seen. On the benefits of manual driving as discussed in this paper (paragraph 3.1) Sheller states "… a conflict between an ethics which is concerned with aggregate effects of personal action on the world at large and a morality that sees caring in terms of more immediate concerns such as one's partner and children" (Sheller, 2004, p. 229). I believe Sheller, who writes more on the sociological aspects of vehicles than the moral side of the debate, touches upon the subject of conflicting reasons. Looking at this statement we can imagine an example in which an agents has to choose between buying a more expensive autonomous vehicle with their entire savings, or buying a cheaper second-hand alternative and keep a part of their savings intact, allowing them to do a number of things which otherwise would have been impossible. Having established that using an AV is morally preferable to using manual vehicles, the morality of the reasons for choosing an AV seems to outweigh the alternatives. However, we would not consider the reasons to create a morally requiring situation for the agent in choosing an AV, either because the reason lacks moral requiring strength or that there is a nonmoral reason to perform an alternative act. Alternatively, if the agent were to choose an AV at the cost of their personal (financial) luxury, safety or freedom, we would consider this morally praiseworthy. Furthermore, it can even be the case that the agent has more moral reason to perform this act than any available alternative, which is one of Portmore's criteria for supererogatory acts.

Moreover, if we re-examine Urmson's examples and recognition of acts which can be considered to be heroic or saintly (as discussed in paragraph 4.1.2.), it is important to note that these definitions are merely used as examples of supererogatory acts, not as the necessary characteristics or criteria one could use to classify any action as a heroic act. That is to say, choosing an AV does not involve similar consequences to choosing to jump on a grenade but for some there will be severe consequences nonetheless. Having shown that the act of choosing to use an AV cannot be considered to be a moral obligation and having further established that there are both moral and nonmoral reasons against switching to AVs (and subsequently that nonmoral reasons can be morally overriding) I conclude that for some drivers switching to AVs can be considered to be a supererogatory act.

5. Conclusion

In this chapter, a brief reconstruction of the structure of this paper will be given. After which, a number of final notes and question for further deliberation will be left to the reader.

5.1 Summary

In this paper, an overview was first created of what is generally accepted as 'the benefits' of AVs, and the general direction of developments surrounding AVs. Furthermore, an assumption was made based on this, stating that in the foreseeable future financial reasons would no more be an argument against the usage of AVs than they currently are for manual vehicles. The benefits of AVs as described were then philosophically interpreted to investigate if the consequences of switching to AVs were really desirable and could be considered to be morally best.

To establish this, theories were introduced by a number of writers. The aim here was not to find and explain theories that could serve as summaries of their respective fields of research, nor were they necessarily the best fit the subject of AVs out of all the writing in the field. Instead, these writers and their work serve as a representation of the philosophical discourse and are used to introduce and 'test' the assumptions surrounding AVs. Firstly, the moral area of interest was shown to have moved in choosing this topic. Whereas most moral discussions on AVs are cantered around the choices the car makes and similar discussions on manual driving likewise focus on the choice of the driver, using the work of John Harris, a case was made that this has now shifted to the driver's choice on whether or not to *use* an AV. This means that there is still an option of making the 'wrong' choice, but instead the choice is no longer made while in the car.

Secondly, a general theory of consequentialism was introduced, which was explained to mean that the rightness or wrongness of an act is determined foremost by its consequences. Furthermore, a more specific consequentialist theory was introduced in the form Mill's utilitarianism. This theory proposes that the right and wrong of actions are indeed determined by its consequences, but specifically by whether or not they

produce happiness or do not produce the opposite of happiness (which is further defined by Mill as pleasure or absence of pain). When these theories were applied to the case of AVs two thing became apparent; 1) general consequentialism would consider the predicted/possible decrease in deadly accidents that may result from switching to AVs a moral benefit, and 2) utilitarianism would agree but shows us that it is not necessarily as straight forward. To elaborate, an example was given showing that an expected outcome of switching to AVs is not only a decrease in fatal car accidents but also negatively impacts the joy that is to be had from driving manually. Since Mills theory focuses on pleasure a case could be made for sticking with manual driving as it does precisely that. However, Mill also remarks that we have the ability to distinguish between different pleasures based on our experience, ultimately leading AVs to be the better choice.

A similar approach was taken in discussing deontological theories. A general description of this theory was given to show its focus on duties that should always be fulfilled, no matter the consequences. One specific theory was further explained to give a better understanding of this, namely that of W.D. Ross. In his work, Ross defines eight prima facie duties. He explains that these duties should not be seen as duties themselves, but more as characteristics an act has which in favour of performing that act for as long as it is not trumped by a proper duty. Examples were then given showing how the act of choosing an AV would the characteristics as described in these eight prima facie duties.

Following this, an analysis of automotive culture was given based on the work of Mimi Sheller, which provided arguments in favour of using manual vehicles instead of transitioning to AVs. Sheller was shown to argue that automotive culture is more than just practical use and style. Instead there is a cultural, social and emotional dimension we do not fully appreciate in this discussion. Subsequently, an explanation was given of four specific dimension Sheller addresses in her work and examples were given on whether or not these spoke in favour of switching to AVs. Two conclusions were drawn, specifically that 1) although there are benefits to manual driving and transitioning to AVs would come at a certain cost, this will only impact a section of drivers that thoroughly enjoy driving, 2) instead all other drivers will experience changes as a consequence of transition to AVs in society in such a way that they will impact their lives so much that is may even change who they are as a person.

It was determined what the benefits of AVs are, and furthermore having determined that these benefits can be considered as benefits from a philosophical standpoint as well. However, a case was also made that there are a number of reasons why some may want to stay with manual vehicles, and that switching to AVs for some would come at a cost. Having established this, the question was raised whether or not switching towards AVs could be considered a moral obligation in respect to the expected benefits or even a supererogatory act in respect to the sacrifice.

Firstly, Darwall was used to introduce and explain what moral obligation is taken to mean. Darwall's theory on accountability was explained and the way moral obligations are the demands that equal, rational agents impose on each other was explained. This was furthermore used as an introduction to a more specific theory on moral obligation, namely that of Scanlon. Scanlon's contractualism was explained and shown to be in agreement with points from both Ross and Sheller's work in regard to conflicting reasons. Scanlon's theory on the general acceptance and rejectability of principles of reason was then explained and used to conclude that because a minority of people may have grounds to reasonably reject or be adversely affected by any such principle, the usage of AVs is not considered to be a moral obligation.

Lastly, the concept of supererogation was discussed. These types of actions, which are generally considered actions that go beyond the call of duty, were further explained by using Urmson's work. Urmson described three types of actions that are generally accepted in ethical theories but are shown to be insufficient to categorize all actions. A different type of supererogatory action was introduced, namely those which we would call heroic or saintly. Subsequently, Portmore was introduced and an explanation was given on the conflicting reasons agent may have, whether they are moral or nonmoral. Portmore's theory on the strength of these reasons and justification was further discussed, showing that nonmoral reasons can be just as important as moral reasons in determining the morality of an action. Given his definition a conclusion was drawn stating that switching to AVs would be considered a supererogatory act for some agents.

Having established the practical and moral arguments for and against the usage of AVs, and having shown the complexity of the decision along with the implications of a full-scale transition may have on our society, I believe this does bring us to the second conclusion of this paper. Simply stated, we as a global automotive society are focussing too much on the expected benefits of AVs, and are not considerate enough of the variety of implications this development may bring to society and the people in it. I would argue this means we cannot simply choose to switch to AVs completely, or even be obligated by any rule or law, ethical believes or moral convictions to do so as it would not match our current automotive emotional experience. To adapt and acquire these new 'matching' affective relations and feel this new embeddedness of the car with our lives and of ourselves with the car, will require a long transition period for all aspects involved. While some may consider it their duty to transition to AVs and others may deem it the greatest of sacrifices if they had to give up manual driving, it could not be considered a moral obligation and would be a supererogatory act. It is however, obvious that this development, which seems unavoidable, will impact our society greatly, and will even impact those who do not drive.

5.2 Alternative questions

In the course of this paper a number of unexplored questions and subjects have been touched upon but set aside as they fall outside the scope of this work. I would like to point out a number of these, as I believe they are most interesting for further research. Firstly, an example was given in which two countries that are heavily reliant on each other for international trade have a vastly different level of automation. The question raised here is twofold, namely how would this disparity influence international development and transition towards fully automated roads and which authority can act in what capacity to manage this intricate interaction?

Secondly, a practical solution for those who wish to continue manual driving is introduced. Namely, if an agent drives solely for the joy of driving he or she can also do this on a closed circuit shared exclusively with other who share this passion. This would turn manual driving into a sport of sorts, not unlike racing cars or fighting a boxing match currently is. This however raises a the question of whether we should consider acts which have generally forbidden in our society (fighting, driving extremely fast, or

possibly driving manually) and allow them to function as entertainment knowing they are dangerous.

This also brings me to the last question I would like to leave the reader with. The argument in this paper is that switching to AVs is a supererogatory act in our society. This society is however, already a society in which we are allowed to do a number of things, which we know might not be the best actions as discussed in the previous paragraph. A simple analogy can be made to show how this may be applicable in the case of AVs. For instance, we can accept that not switching to AVs brings pleasure but saves lives, but we also accept that buying nice clothes instead of donating money to charity brings pleasure but costs lives. Subsequently, we live in a society in which it is acceptable to buy nice clothes, in which case it should also be acceptable to keep driving manually. This argumentation here is not logically sound but is not indented to be so. Instead the intention is to raise the following question. If we lived in a morally more demanding society, would the theory discussed in this paper still hold true? Moreover, try imagining what our society would look like if it was more morally demanding in general.

6. Bibliography

Anderson, J., Kalra, N., Stanley, K., Sorensen, P., Samaras, C., & Oluwatola, O. (2016). *Autonomous Vehicle Technology.* RAND Corporation. Santa Monica: RAND Corporation.

Arbib, J., & Seba, T. (2017). *Disruption, Implications and Choices.* RethinkX.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science , 35*, 2, 3.

Darwall, S. (2007). Moral Obligation and Accountability. In S. Darwall, *Oxford Studies in Metaethics, Volume 2* (pp. 91-118). Oxford: Oxford University Press.

Davies, A. (2017, April 4). *Mercedes Promises Self-driving Taxis in Just Three Years*. Retrieved November 12, 2017 from Wired: https://www.wired.com/2017/04/mercedes-promises-self-driving-taxis-just-three-years/

Forrest, A., & Konca, M. (2007). *Autonomous Cars and Society.* Department of Social Science and Policy Studies. Worcester: Worcester Polytechnic Institute.

Giarratana, C. (2016, December 6). *Vehicle-To-Vehicle Communication Systems*. Retrieved September 7, 2017 from Safety Resource Center: https://www.trafficsafetystore.com/blog/vehicle-to-vehicle-communication/

Harris, J. (2011). Moral Enhancement and Freedom. *Bioethics* (25), 102-111.

Hawkins, A. (2017, May 9). *South Korea says it's building the world's largest test site for self-driving cars*. Retrieved November 2, 2017 from The Verge: https://www.theverge.com/2017/5/9/15596366/south-korea-self-driving-car-test-site-worlds-biggest

Mill, J. (1979). In *Utilitarianism* (pp. 14, 16, 17, 19). Indianapolis: Hackett Pub. Co.

Portmore, D. (2008). Are Moral Reasons Morally Overriding? *Ethic Theory Moral Prac* (11), 369, 373, 380.

Ross, W. (1930). In *The Right and The Good* (pp. 17, 20, 21, 32). Oxford: CLARENDON PRESS.

Scanlon, T. (1998). In *What We Owe to Each Other* (pp. 203 - 205). Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Sheller, M. (2004). Automotive Emotions. *Theory, Culture & Society , 21*, 221- 225, 227, 229, 234, 235, 237.

Sinnott-Armstrong, W. (2015, October 22). *Consequentialism*. (E. N. Zalta, Editor) Retrieved November 25, 2017 from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/consequentialism/

Stewart, J. (2017, April 25). *Google's Finally Offering Rides in its Self-driving Minivans*. Retrieved November 6, 2017 from Wired: https://www.wired.com/2017/04/googles-finally-offering-rides-self-driving-minivans/

Urmson, J. (1958). Saints and Heroes. In *Essays in Moral philisophy* (pp. 62, 63). Washington: University of Washington Press.