

Bachelorscriptie Kunstmatige intelligentie

The best course of action: on the wicked problem of self-driving cars

Student	Sietze Kai Kuilman
Begeleider	Jan Broersen
ECTS	7,5
Datum	14/01/2018

Table of Contents

Abstract.....	3
Introduction.....	3
SDCs as a wicked problem	3
Coping strategies for wicked problems	7
Navigating towards an implementation of SDCs.....	7
Discussion.....	9
Conclusion.....	9
References	10

Abstract

Self-driving cars could save countless lives. Implementation of such technology seems desirable however how can we do it well? Other articles mostly look at one aspect of the problem either through a legislative aspect, social, or an ethical aspect. This paper looks at the system as a whole. Implementation of self-driving cars seems to have the characteristics of a wicked problem. Wicked problems are social problems that are hard to solve because they have interested parties which are often fundamentally at odds with each other, solution depends on the way the problem is formulated and are impossible to implement on a trial and error basis. In order to implement a solution we need a coping strategy. Three possible strategies are discussed in this paper: authoritative, collaborative, and competitive. While an authoritative strategy might allow for a quick implementation and a collaborative strategy for a nuanced and well integrated implementation, a competitive strategy would allow all stakeholders to have a say in the implementation without sacrificing too much implementation speed.

Keywords: Self-driving cars; Social planning; Wicked problem; implementation; coping strategies.

Introduction

Self-driving cars are becoming more and more of a reality. While the driver takes the backseat, the car can drive itself towards the destination. There are advantages to such an innovation. A self-driving car won't fall asleep behind the wheel. It won't even need breaks to rest. But perhaps the most important difference self-driving cars will make is around safety. As it stands about 90% of traffic accidents are due to human error (Gao et al. 2014). Autonomous vehicles could lower this percentage which is, in any form or shape, a welcome addition (though we also need to consider new problems like malfunctions and cyberterrorism). Even so there will still be accidents as some situations cannot be avoided. Self-driving cars will enter a huge and complex system with not only other self-driving agents but humans and animals too, some of which might have unpredictable behaviour. A man suddenly jumping in front of a vehicle that is speeding on the highway is something that could happen and in some of these cases the only options the car has is to swerve and kill an innocent bystander, to hit the man jumping in front of the car. A lot of questions arise with these kinds of situations. How does the car behave and who is to blame when casualties are involved? These decisions on behaviour and where to place blame are of importance to everyone if we are going to let self-driving cars loose on the road. However can we implement self-driving cars in such a way that everyone is satisfied?

The research on self-driving cars and their ethical design is a topic of hot debate. The problem is becoming more and more important as the future of self-driving cars is drawing nearer. Many facets of this problem have been discussed. Who wants to ride in a self-driving car and what do people find acceptable behaviour? (Bonnefon et al. 2016). Who is to blame when a crash happens and what legal rules should

be in place to make sure the blame is placed there? (Belay 2015) Of course the question remains how we can actually design an ethical code for a car without going over boundaries (Lin 2016). All these questions are of concern as the integration of this new technology is desirable.

How we best implement self-driving cars, into society? Can we actually implement them in a "best" way? The purpose of this paper is to examine the problem of implementation around self-driving cars through the lens of wicked problems and then to discuss how we might go about coping with it.

SDCs as a wicked problem

The problem of self-driving cars, after this referred to as SDCs, is often framed through the ethical behaviour of the vehicle itself. We observe how a vehicle would act in general cases and then move into more and more fringe cases, for example the classic trolley problem wherein we need to choose between killing off one individual or five individuals (Thomson 1976). Looking at the problem this way means we are mostly limited to ethical implications and view those through a case by case basis. This means that we put a lot of attention towards the fringes of the ethical behaviour of a car. While it may be true that in some off chance SDCs will choose to ride over someone for all the wrong reasons, in general SDCs will perform better than any human driver whatever ethical design might hide underneath the hood (Gao et al. 2014). So why haven't we implemented them? It seems only logical to implement this technology as soon as possible. Safer cars means less road kills. However to view the implementation of SDCs in such a way would be to oversimplify it. Implementation of such a technology will have long lasting consequences. Of course we need to consider what kind of general ethical behaviour the car has. Do we value community, where one might be sacrificed for the group? Or is self-preservation key? There is also the question of responsibility. Who do we blame when someone is killed? Can we actually blame someone and if so on what grounds? Can we let children drive these cars? Do we still want the option to take control over the wheel? We also need to take into account how our behaviour changes if we can always drive without worry. Will the environment suffer because of our increased traffic? Do we need to ban human drivers off the roads? Can we make sure everyone actually wants to drive such a car? What level of automation is best for an SDC? Giving the car total control or letting it act more like an assistant? Some parties might not want to lose control over the wheel as they might say it would impede their freedom. While others might argue that it could be safer if we give up control of our vehicles all together.

All these questions arise because we have wildly different opinions when it comes to planning and designing a society. Because of this the implementation of SDCs seems to fit the bill of a wicked problem.

The distinction between tame problems and wicked problems was first proposed by Rittel and Webber (Rittel Webber 1973). It is a distinction to clarify why certain problems are hard to solve, in the case of social planning. Tame problems, which are the opposite of wicked problems, are well defined problems, which have clear solutions. Consider the problem of solving an equation. We can use well-defined operations on the equation to eventually come to a conclusion. In contrast wicked problems have none of these traits. They are in the social sphere, where people often are at odds on how to design or implement something. Tax rates are such an example, where more socialist standpoints are often viewed through the eye of the collective (making use of redistribution) and on the other end of the spectrum where we have libertarians who would argue that tax equals theft. It seems unlikely that there would ever be an implementation to accommodate both parties. According to Rittel and Webber there are ten markers to see whether or not a problem is a wicked one (all the markers named below are cited from the original article). I intend to explain them all and compare them to the problem of SDCs.

1. There is no definite formulation of a wicked problem.

Rittel and Webber argue that the information needed to understand a wicked problem depends upon its solutions. What is meant by this is that the some problems for example poverty can depend on many different factors. If one proposes that poverty is partly dependant on low income and this is caused by faults in an educational system then the problem of poverty suddenly has ties to the gaps in an educational system. However if one decides that poverty is also caused by problems in mental or physical health then we need to add our medical care into the equation. We could also determine that poverty is caused by spatial dislocation and look at geographical data instead. Because of this we can continually add information and based on the data we can propose possible solutions. As such the solution and the formulation of a wicked problem are linked.

For SDCs we can argue that this link is clearly there. Implementing SDCs is invariable linked to a stance on ethics. We could argue that self-preservation is key because we have a biological urge for survival, to come forth with such a hypothesis we would probably include different studies on human behaviour to support this. However we could also argue from a societal standpoint that it would be better to think about the community. Having cars that will always protect the driver might lead to some unwanted fringe cases. Those cases might be bad enough to say that an implementation skewed towards the community might be more beneficial.

To say that the implementation of SDCs is purely an ethical problem however is to cut it short. We also need to think about the best way to implement it on the basis of environment and safety. What rules would govern a SDC? All of these angles to look at the problem can have various solutions depending on what is more important to the

implementer. There are various parties involved for example the state, manufacturers, and individuals. All of these parties will want various things when it comes to the implementation. Along with those parties come various solutions that are always more in favour of the party involved. A manufacturer will benefit from cheap and perhaps polluting cars. Environmentalists however might skew towards a more climate friendly solution, perhaps even arguing that all SDCs should be powered by batteries because the increased traffic on the road might have serious consequences on the environment. In short any solution proposed for the implementation of SDCs will be based on a certain formulation but that will not be a definite formulation. While I could argue that the set of possible formulations for SDCs is limited. One could always go on a limb and state that problem X of SDCS (for example traffic accidents) is caused by Y (say education). The entire point is that we haven't postulated all possible causes. The more data we will collect the smaller the set of possible formulations will be.

2. Wicked problems have no stopping rule

When searching for a solution to a wicked problem one never has to stop searching. This fact can be deduced from the first marker Rittel and Webber propose. As a wicked problem has no definite formulation we can always consider more variables, to include for our possible solution. As such there is no reason to say the problem is solved.

In the case of SDCs we can think of new ethical designs we might not have considered or other factors to include in our design problem. As such we can always gather more possible information on whether or not a certain implementation will be good. Moreover context might change and so will technology. With the rise of new technology we might be able to implement SDCs in a radically different way. Of course while we are adding new possibilities we are also losing others. In the case of SDCs some might want to argue that the quicker they are on the road the better. While we might not have a clear stopping point, since the solution still depends on what you value, there are points to be made to make sure that a search for the implementation of SDCs isn't just an eternal search without any solution.

3. Solutions are not true-or-false, but good-or-bad

According to Rittel and Webber there will not be a true-or-false kind of solution in the case of wicked problems. We have various interested parties who all have their own solutions and opinions. Those solutions may vary because of personal interest, different value structures or ideological preferences. As such a proposed solution to a wicked problem will more likely be satisfying or good enough.

I have named some of the parties that are involved with SDCs. We have manufacturers, the state, and individuals. These parties have different wants and needs but also within these parties ideas about implementation vary. The amount of freedom a driver is given is such an example, which in

turn has to do with the level of automation. Again we can think of individuals who would argue for the rights of individual liberty, while others might argue for the safety of the collective. As long as we remain ideologically at odds with each other, which might not be bad thing at all, a true or false solution seems out of the question.

4. There is no immediate and no ultimate test of a solution to a wicked problem

As with most societal problems, a wicked problem can have long lasting consequences. Implementation of taxes in a certain way could lead to poverty which in turn could lead to an enormous amount of possible complications in our society, for example riots. The consequences of designing a society in a certain way will have consequences with a compounding effect, as it is the stone thrown into the water that creates the ripples.

This is most crucial to think about when implementing SDCs, as the design of SDCs can eventually lead to radical outcomes. It would be most easy to point out that having code that specifically targets part of society, in the case of choices between life and death will have societal consequences. Again if we grasp at the straws of freedom then we could wonder what would happen if people do not need driver licenses anymore. What will the impact be when we let children on the road because the cars can drive them anywhere? Of course most social changes will have mild effects but what about the lives of cabdrivers and bus drivers? What impact will self-driving public transport have on their lives? The fact is not that we cannot find an ultimate test but that we will most often have missing or uncertain data, which would mean that we haven't looked at the problem from all angles and in turn could mean that our test might not encompass everything. While the consequences can certainly be modelled, we should wonder what data we are missing.

5. Every solution to a wicked problem is a “one-shot operation”; because there is no opportunity to learn by trial-and-error, every attempt counts significantly

Rittel and Webber argue that wicked problems leave traces that cannot be undone. We cannot build a freeway to see whether it is satisfying or not. If we would build a freeway then we couldn't easily reverse the process and if it happened to be unsatisfying lives would be changed. Not only would it cost heaps of money to reverse the act, it would also change the landscape. In other cases like the particular design of a city it might have lasting consequences on its inhabitants.

If we would implement SDCs for the sake of trial and error then we are talking about a high speed vehicle with passengers and pedestrians surrounding it doing something hardly tested. Although it might be true that we can update the software of SDCs with more ease than we can remove a

freeway from our landscape. There are other things to consider. Implementing SDCs in anyway will have consequences that count significantly. The best way to explain it is by pointing at nuclear energy. This kind of energy is now safer and creates little waste but despite this fact people are cautious because they remember it differently. If we implement SDCs on a trial-and-error basis we might just end up distorting the general view on SDC, which could mean we won't implement SDCs at all.

However this only applies if we look at the long lasting consequences and on a most abstract level and if the implementation is a form of all or nothing. We can start adding cars to the environment as is. The change would be gradual. Or we might be able to implement it in simulated environments, have people watch the car, and question whether its behaviour is correct. Besides most manufacturers are implementing SDCs exactly through the use of trial-and-error and there is still a driver present in such tests. So I have my doubts with this marker.

6. Wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan

If we accept that wicked problems have no stopping rule and that it is problematic to formulate the problem because we can always add in more factors. Then we can also accept that we cannot know whether or not we have exhausted every possible angle. As such the set of potential solutions cannot be described.

Considering the fact that the implementation of SDCs has long lasting consequences on society and driver behaviour, I would argue that it is impossible to exhaustively describe all possible implementations. There are simply far too many factors that compound over time to know whether we have chosen the best implementation. However as I said before the more data we collect the better we will know which factors to look at and which to ignore. While wicked problems might not be exhaustively describable we might find a set of solutions which are adequate and since we can update the nature of cars we can implement an adequate solution first and adapt it in such a way that it incorporates more data each time.

7. Every wicked problem is essentially unique

Rittel and Webber argue that when one does solve a wicked problem you won't have solved problems similar to it. There is no set or class of wicked problems which allows you to abstract certain ideas and solve all of them. This again comes back to the fact that wicked problems are contextual. What works in one place might not work in another because the timing is different or the people are different.

The technology of SDCs isn't similar to anything else. How people think about cars and SDCs is specific and only a handful of specific companies are invested in the project. The problems of SDCs could be compared to the problem of

vehicles in the early twentieth century. The replacement of horse and carriage was at the time something we rebelled against. People were afraid of what it would do to society. However our society has sped up considerably since those days. Context makes wicked problems unique and the implementation of SDCs seems context dependant. Of course on some abstract level a lot of the problems surrounding new technology will be the same, which is an indicator that some factors will be identical. We can look at these problems to perhaps prevent certain pitfalls. So yes technically the problem is unique but it might have a lot of similarities with comparable problems.

8. Every wicked problem can be considered to be a symptom of another problem

According to Rittel and Webber a problem can be seen as a symptom of a more broad and general problem. For example the case of criminality might have to do with moral decay or deficient opportunity, or any other explanation you wish to impose. However such moral decay for example might only be a symptom of a broader problem. Perhaps it is the way we raise our kids or the way we educate them. The level at which a wicked problem is settled is arbitrary. On one hand we wish to solve a problem at the highest level as it would solve the problem at the roots but on the other hand the problem becomes broader and broader the further we ride up the tree. A broader or more general problem becomes harder to solve.

The implementation SDCs could be seen as a symptom to reduce reckless behaviour on the roads. This in turn has its roots in behaviour and when it comes to society and behaviour we can choose various arguments on how to solve it, which in turn is a problem of how we wish to design our society. The implementation of SDCs could also be seen as symptom to reduce traffic. Traffic can exist because of many reasons, one of which is of course gawking at crashes. Again such a problem can be lifted towards another level. Of course it is merely a point of definition, which Rittel and Webber continually return to. As long as we accept that the definition of SDCs is indefinite then it we can logically conclude that such problems as these follow.

9. The existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution

It is clear looking at marker one, six, and eight that we can try to define a wicked problem in any way we wish and that it also determines what solution we might get. The question remains what choice is right? Do we say criminality is a symptom of moral decay or deficient opportunity? Either one seems viable. There is no way to determine the correct explanation. Of course we can think of what is most plausible. But plausibility is still determined by the information one has, which can always be lacking.

On the subject of implementation of SDCs we need to consider what is most important. Is it important to save

lives? Is it important that we get fringe cases right? Or is it important that the technology is implemented as quickly as possible? We can also consider why we want to implement SDCs and what that means. Don't forget that also our personal interest come along with this implementation. If you believe in freedom and self-preservation then implementation of SDCs will differ from someone who is intent on helping the community. Those differences can also be symptoms of something larger. Of course we can create models on what saves more lives, what will perform better. However can we also take into account the individuals happiness and willingness to drive such a car?

10. The planner has no right to be wrong

Wicked problems have their origins in social planning. The last marker includes the planner, he who tries to solve the wicked problem. Rittel and Webber argue that certain principles of science do not apply here. The kind of wrong we are talking about here isn't miscalculations. It's the thesis of Newtonian gravity versus general relativity. In science this is fine. However in social planning we cannot offer a hypothesis and implement it while we wait for a refutation. The scientific method, where formulate a hypothesis and look for evidence against it, isn't enough. While we wouldn't want scientists to be wrong, their theories will allow us to better understand the world. When a theory is at an end we can postulate another. And when a hypothesis is refuted the original author isn't blamed for postulating the hypothesis. The implementation of a planner's hypothesis however has serious consequences on the world. Thus Rittel and Webber state that the planner would be liable for the consequences.

When we look at SDCs we can immediately wonder who would be responsible for the implementation. However the point Rittel and Webber makes is clear. We cannot be scientists who try a certain hypothesis just to see whether there is any evidence against it, the consequences would be too grave if said hypothesis is later refuted. On the other hand we do have more leeway than most planners. We aren't creating a landmark or sculpting a cityscape. We are putting machines on the road which can be updated and adapted. While they could have long lasting consequences they are more easily reversible.

The implementation of SDCs has some of the markers of a wicked problem. What Rittel and Webber demonstrate is the fact that definition is hard and that various parties have varying ideas about what is most important. These two factors seem to fit the implementation SDCs. I would argue that it is enough of a basis to call it a wicked problem, if only to see whether or not social planning might offer us some solution to the problem. Considering this we must accept that we will not find a true-or-false solution here. However a solution doesn't have to be perfect. It has to be satisfying.

Coping strategies for wicked problems

While solving a wicked problem may not be doable we can find ways to cope with them. (Roberts 2000) Roberts describes three generic coping strategies to deal with wicked problems: authoritative, collaborative, and competitive. These strategies are generic because each wicked problem is unique and will require a case by case formulation. Neither will these strategies offer clear or concise solutions as it may be clear that such a solution will not be possible anyways. I will explain these three strategies here and in the next section I will apply them to the implementation SDCs.

Authoritative

An Authoritative strategy in the case of wicked problems means that one or a small group of stakeholders gain the authority to define the problem and come up with a solution. The basis on which these stakeholders are chosen could be hierarchy, knowledge and expertise, or whatever else is applicable to the situation. Other parties must transfer their power to this select group and abide by their decision. In short power is divided unequally to make the decision process more manageable. The advantage to such a coping strategy is the simplification of the problem. It is hard to decide on something if numerous parties want to be part of the decision making process. Besides the parties chosen to represent us in the matter could make problem solving more professional. There are more tools available to some experts, especially those who have intimate knowledge about the subject. They may know certain procedures or tools which might save time and resources.

However experts and authoritative figures can be wrong in what they prioritize. In the case of wicked problems this could mean they are wrong about the formulation of what the problem actually is. In such cases this can also mean that decisions are made too hasty and if the parties involved think from a different frame then citizens might suffer for it. Leaving problems solely to experts could mean that citizens interest themselves less and less in the problems which influence them as they have no say in the matter anyways.

Collaborative

Collaboration is the premise here. Various parties join hands to collaborate on the problem in hopes of finding a possible solution. It is based on compromise but in such a way that all parties gain more than they would have in their worst case. Of course the increase in the number of stakeholders means that more voices will be heard, more experts are on team. This means a more nuanced view of the problem. Besides it can also mean that various parties can share the costs and the benefits. Think of competitors working together with a supplier to improve their services. In this case the competitors both benefit in cost while redundancies are eliminated. It offers a chance to let all parties focus on what they do best.

On the other hand adding a new stakeholder to the bunch also means we have to add a certain cost. More parties mean more meetings and more people who we need to get on our

side. Moreover collaboration is a skill, something which needs to be acquired. If one or more parties lack this skill then collaboration can end poorly and even if such collaboration does work out it does not guarantee that the outcome will be satisfactory.

Competitive

Competitive coping strategies are based on a 'zero-sum game'. If my opponents win, they get to define the problem and choose the solution, I lose. There is no compromise involved. Winner takes all. This is different from authoritative as the deciding parties aren't chosen beforehand. It is a struggle for power. Depending on the problem of course you might need to build a larger following or position yourself in such a way that you can easily take out the competition. In the long run competitive strategies can move into authoritative strategies. The advantages of a competitive strategy are plain. Competition is often crucial for innovation. New ideas are often reached in the market sector because we want to increase profits and as long as no new way to tackle the problem is found and the conflict is great we can wait till the dust is settled. It might be preferable to stay put instead of moving forward in any direction. Besides competitive strategies offer one other major benefit, it keeps power circulating among the competition as one day you win and the next day I might win.

However competition can mean stalemates. If competitors have enough power to stop others but not enough to actually do something about the problem then we might end up doing nothing. Competition also consumes resources which otherwise could have been spent on problem solving.

Navigating towards an implementation of SDCs

How does this apply to our wicked problem, the implementation of SDCs? First we must define our stakeholders, the parties with a vested interest in the matter. Belay proposes four parties who have legislative interest in the subject of SDCs, while the implementation of SDCs is more than a legislative procedure I would say that it does cover the parties who have an interest in the matter (Belay 2015). First off we have the individual. Meaning those who drive cars but also those who come into contact with traffic, have a clear interest in the case as they will use the product. Secondly the manufacturers, any company that currently produces cars and any company that wishes to produce SDCs, want to implement the technology or keep others from implementing it. Thirdly the state is the party that is invested in roads, has safety guidelines, and is also concerned with certain ethical decisions. The problem of SDCs could be one of those decisions. Lastly Belay includes the insurers, which in this case have a vested interest in the matter because their profits are built upon a system that might be replaced. In the case of implementation of SDCs there is one more party that has certain influence on the

matter and a clear interest and that is the oil industry. It is important to note here that these stakeholder groups can still be split into smaller fractions as individuals will differ on what is ethical and what we should value. The same goes for manufacturers, states, insurers, and the oil industry. State might differ for example in their needs because road signs or rules are different. I do not intend to divide them specifically as that could be a paper all on its own.

Second we must define what matters. Since this is a wicked problem I will not be able to describe every angle on the problem. I can only describe the things that I figure are crucial to the implementation, namely speed, ethical behaviour, level of automation, responsibility, and costs. Implementation speed equals the rate at which SDCs will be on the road. The faster SDCs are out there the more lives we save. The ethical behaviour of the cars itself is also of importance to how well it is implemented though I would argue that whether or not the implementation is good largely depends on your views of on ethics, which is a necessary inclusion but hard to measure. We also need to consider what level of automation we are talking about. Is this a full blown automaton or merely some kind of driver assistant? This has a lot of social implications, from the lives of cab drivers to getting your driver's license. Responsibility hinges together with implementation speed in some sense because clarity surrounding it might increase the chances of getting SDCs on the roads. However since responsibility is crucial to the problem of SDCs, as they are (to varying degrees) autonomous yet blameless, we need to figure it out in our implementation.

There is a lot of research on the costs and benefits of SDCs. Safer cars might lead to fewer crashes as I stated in the introduction. However we need to consider there are other risks like malfunctions, cyberterrorism (Bilger 2013) and offsetting behaviour (taking additional risks because participants feel safer). There could be reduced seatbelt use and pedestrians who become less cautious and responsible around SDCs. (Milliard-Bal, 2016) SDCs will also affect public transport, parking costs, commuting strategies, roadway design, and the environment. The different costs and benefits are well presented in a paper by Litman. (Litman 2014)

Authoritative Strategy

Which stakeholders should be given power? From the possible choices, it seems unlikely that we can give individuals the power as they are far too large a group to make collective decisions. Insurers and Oil companies might have incentives that could prolong implementation. That leaves us with Manufacturers and the state. Of course what choices would be made by such a stakeholder are completely open to interpretation. One such interpretation or perhaps better said proposal is that of Belay who argues for a state implemented SDCs ethics of self-preservation. This means that the state would draft up legislature that would define under what parameters SDCs must be built. Another interpretation might be a utilitarian approach (Bentham

1798). A kind of approach where there is one greatest good and a clear ethical solution that can be calculated.

A serious advantage to this strategy is clarity. All stakeholders will know what to expect surrounding themes like ethical behaviour and responsibility. There would be guidelines on what happens in the case of an accident. Besides the state can define the levels of automation and create laws accordingly. Not to mention that such clarity and quick decision-making by one stakeholder could lead to a quick implementation of SDCs. Manufacturers might be less hesitant if they know the risks involved. Another advantage is the amount of recourses and data a state has access to. They might be the more capable stakeholder to assess the costs surrounding increased traffic, pollution and safety.

The disadvantage comes in the form restrictions. How strict of an authoritative strategy will this be? Will it be different between states? If so then driving through different countries might become very difficult. How about disagreements? Will we still have public transit once these cars hit the road? This only applies if we are talking about the highest level of automation. However if this is the case then what happens to people who disagree with the states design? Perhaps the most important question to ask is will it be a singular ethical code or will it be multiple? That entirely depends on the states outlook. If a singular code does arise, that makes active choices and does not determine behaviour on dice rolls, then how do we make sure the consequences aren't too severe? Lin mentions that such a singular code could have disastrous effects on our behaviour. For example think of a car that has to choose between two motorcyclists, one with a helmet and one without. The car will hit motorcyclist who wears a helmet, because they have a better chance of survival. However this means it will be safer to leave your helmet at home. Such behaviour might be avoided if we have multiple designs on the road.

Collaborative Strategy

When we consider a collaborative strategy we need to ask the following question: Who do we invite to the table? While we have certain parties who are invested in the current state of affairs, namely the insurers and the oil industry, we might not want them at the table as they might slow the process down. The following party that would be difficult to give access to the negotiations is the individuals, though if we could then something like contractarianism would be the way to go (Rawls 1971) (Hobbes 1651). Of course we could have something akin to representatives, though ideally the state should represent the individuals. As such it might be more efficient to have collaboration between manufacturers and states.

The advantage to this strategy is a more nuanced approach. If these stakeholders collaborate then they will have different wants. The state might advise on certain restrictions to prevent harm to their citizens. Manufacturers

on the other hand can advise on ethical design and or car design. Together they could create a solution which is bound by certain limits but free enough to deliver multiple ethical designs, taking out perhaps more radical approaches to SDCs (say a car that specifically targets on age or any other discriminatory variable). This would also take away some of the more difficult decisions from the state namely ethical design and leave that up to the manufacturers. Since the state can provide legislative clarity, the manufactures would know the risks involved. Not to mention the fact that the state can argue in the name of safety that all ethical designs can anticipate on one another.

Of course the disadvantage to this strategy comes in two forms. Discussion between state and manufacturers might take long which would lead to a low implementation speed. The discussion could also lead to an interaction between state and manufacturer that isn't favorable towards individuals. We have to take into account that crony capitalism is a possibility.

Competitive Strategy

A competitive strategy in the case of SDCs is best held by proxy, in my opinion. We won't have states designing and producing SDCs anytime soon. So we can be competitive through manufacturers. While they are but one stakeholder this competition would trickle down, as a competition between manufacturers would mean a competition between individuals. They have to buy the product in the end. The state could subsidise one manufacturer to show their preference and so could the oil company or the insurers. The manufacturers would be the platform through which the other stakeholders can show their choice. In short this could resemble something like distributive justice (Nozick 1973). Each individual gets the choice to argue what implementation is best and they would do so through voluntary transaction.

The advantage of this is the mitigation of an authoritative code, as multiple manufacturers would also mean multiple ethical designs. Since this is a great way to differentiate from the competition. Implementation speed will probably be on par with an authoritative strategy as companies are rushing out to be the first on the road. Another advantage is the freedom individuals get. There will be enough choice, hopefully such that everyone can have a car that they wish to have. Of course if one doesn't find such a car and they do deem it important enough then can create it themselves.

The disadvantages come with the exclusion of the state. How can companies be certain they won't be blamed for the mistakes of their car? When is something a malfunction, when is it just a computational decision? The manufacturers might be able to force customers to sign a waiver. However some might not want that and there will still be unclear fringe cases which will mean more risk to manufacturers. Another disadvantage comes with the multitude of ethical designs that could keep on expanding. If the state doesn't set any restrictions then there is nothing restricting companies to create design upon design nor is there

anything restricting them from having radical and approaches to ethical design. There is also nothing that forces them to interact with other ethical designs. The market might be a force that can counteract this but that would mean people would have to be willing to invest time and effort into the subject.

Discussion

Considering the three strategies we discussed above I would argue for something along the lines of a mostly competitive strategy with some authoritative outlines. In other words let the state design some basic limitations (like mandatory interaction between different ethical designs) but afterwards give manufacturers and thus other stakeholders the freedom to decide. This is much along the lines of a collaborative strategy actually but instead of sitting around the table we have two parties working for their own gain while taking away problems from the other. Companies are already vying for the best SDC so we might as well let them as this would ensure an ample set of ethical behaviour. An authoritative layer on top of it might prevent some of the most radical choices or at least provide some clarity surrounding responsibility. This would allow us individuals to support the car we think is best. A competitive strategy does seem the only way to make sure individuals have some say in this matter. Since I categorized SDCs as a wicked problem and the solution to a wicked problem goes hand in hand with the ideology of the researcher. I have to add that I value liberty and responsibility, which is why such a strategy appeals to me. In short we would have a strategy with clarity because of an authoritative outline, a real incentive to have various ethical codes which in turn minimize consequences, and an implementation speed that is quicker than a collaborative strategy. Some safety measures could be enforced through authoritative design.

I included some articles in each strategy as they could provide a basis on which we could further investigate such strategies. That is however far beyond the scope of this paper.

Conclusion

The implementation of SDCs could be seen as a wicked problem. The real point to take away from a wicked problem is the fact that the solution will be bound to some ideological beliefs hold by the researcher. Even if we would simply view one part of the problem of SDCs like ethical design we would still encounter the values held by a researcher. Since the integration of new technology will become a more prevalent question in the future, we might need to consider which of those problems if not all will be wicked. In such a case we will need to establish not only how to solve them but what drives the solution, the values held by the parties involved because those will drive the solution.

Only through discussion and formulation on what the best coping strategy might be will we find a satisfying solution. In the case of SDCs we have a chance to act out not one but

many solutions at the same time thus minimizing their potential consequences in the long run, which might just be the shortest route to our destination.

References

- Gao, Paul, Russel Hensley, and Andreas Zielke. "A road map to the future for the auto industry." *McKinsey Quarterly*, Oct (2014).
- Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan. "The social dilemma of autonomous vehicles." *Science* 352.6293 (2016): 1573-1576.
- Belay, Nick. "Robot Ethics and Self-Driving Cars: How Ethical Determinations in Software Will Require a New Legal Framework." *J. Legal Prof.* 40 (2015): 119.
- Lin, Patrick. "Why ethics matters for autonomous cars." *Autonomous Driving. Springer Berlin Heidelberg*, (2016): 69-85.
- Thomson, Judith Jarvis. "Killing, letting die, and the trolley problem." *The Monist* 59.2 (1976): 204-217.
- Bentham, Jeremy. *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press, 1996 (1798).
- Rawls, John. "A theory of justice." (1971).
- Hobbes, Thomas. *Leviathan*. A&C Black, 2006 (1651).
- Leben, Derek. "A Rawlsian algorithm for autonomous vehicles." *Ethics and Information Technology* 19.2 (2017): 107-115.
- Nozick, Robert. "Distributive justice." *Philosophy & Public Affairs* (1973): 45-126.
- Roberts, Nancy. "Wicked problems and network approaches to resolution." *International public management review* 1.1 (2000): 1-19.
- Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, 4(2), 155-169.
- Bilger, Burkhard. "Auto Correct: Has the self-driving car at last arrived." *The New Yorker* (2013): 96-117.
- Millard-Ball, Adam. "Pedestrians, Autonomous Vehicles, and Cities." *Journal of Planning Education and Research* (2016): 0739456X16675674.
- Litman, Todd. "Autonomous vehicle implementation predictions." *Victoria Transport Policy Institute* 28 (2014).