User evaluations of moral behavior in self-driving cars

Marjolein Zwerver
s4142098
Artificial Intelligence
Radboud University Nijmegen
August 12, 2016
Supervisor: Dr. Pim Haselager
Bachelor's Thesis in Artificial Intelligence

Contents

1. Abstract

In this research project we will investigate how the public thinks a self-driving car should behave when dealing with moral dilemmas that could occur while driving. We will determine to what extent people think possible algorithms for such a car behave in a fair way and whether or not these scores are a good predictor for the willingness to pay for the use of such a car. We will also try to determine the underlying factors for their scores on fairness so they can be used to predict these scores. This is done through the presentation of several scenarios containing some kind of moral conflict based on a number of different factors, followed by survey questions. Using this data, we will provide recommendations for the computational implementation of moral choices in self-driving cars. Finally, we will discuss potential ethical issues surrounding these implementations.

2. Introduction

Research into moral behavior is essential as robots are getting more and more autonomous. They are increasingly involved in completing complex tasks. There are already several domains in which we need to think about moral behavior in robots, such as warfare and healthcare. This will grow in importance as the available technology is advancing.

In the domain of self-driving cars, prominent issues are the debate about a potential need for moral behavior in self-driving cars [1, 2], the resistance against such cars and the lack of technology to implement complex methods. These methods could utilize machine learning and neural network techniques as the technology becomes available [3]. Even though there is still a lot of research to be done on these topics, answering the research questions of this project could be helpful in creating a rational agent that can reason about morality by determining to what degree people would prefer certain algorithms for self-driving cars over others. This will help to improve acceptance of self-driving cars, so more people will drive them in the future. This will, among other things, most likely result in lower amounts of casualties in traffic.

Even though more and more research is done in the field of self-driving cars [4, 5, 6], the best way to implement ethical behavior in such a car is still unclear. Research on utilitarian cars suggests that the utilitarian approach could be a good point to start from [5]. The studies performed in this paper suggest that people are generally comfortable with cars that would minimize the death toll in case of unavoidable harm even, but less so, in cases where the cars occupant was sacrificed in order to save multiple other lives. Unfortunately, as pointed out in [3] this approach has its flaws as it is prone to conflict with our notion of justice. For example: a motorcyclist with a helmet would always be targeted over one without a helmet, because the motorcyclist with a helmet is less likely to sustain injuries. Many people may consider this to be unfair, especially because the motorcyclist with a helmet invested in his safety.

To resolve this issue, several papers suggest a combination of different theories [7, 8]. There is also research on the tunnel problem [6, 9, 10, 11, 12] which is a variation of the trolley

problem. But before we can decide which theories are the most suitable for implementation we must determine what the public thinks about algorithms based on these theories. This is important in order to improve acceptance of self-driving cars which will make use of these algorithms [14]. For this project, it is our goal to get a sense of which decision the passenger of a self-driving car would think is right in terms of fairness and which factors underline this perception of fairness. After all, it is not just the actual fairness of the car's decision in itself, but also the user's perception of fairness that will be influential for the acceptance or rejection of moral self-driving cars. We are also interested in which car people are most likely to spend money on and we will determine if the assigned fairness is a suitable predictor for the willingness to pay for the use of a car in order to generalize our results.

3. Research questions

In order to accomplish our goals, the following research questions were formed:

To what extent do people think certain choices made by self-driving cars in ethical dilemmas are fair?

Which factors influence people's view of fairness of the car's choice the most?

Is fairness as viewed by the participants a good predictor for willingness to pay for use of the self-driving car that makes the preferred choices?

4. Background section

Here we will discuss some background information and references to the main notions and developments that underlie this thesis.

4.1 Trolley problem

The trolley problem is a psychological thought experiment first introduced in 1967 [14] where you have to decide whether or not to flip a switch. If you do nothing a trolley that is out of control will kill five people and if you flip the switch the trolley will be redirected onto a different path and kill only one person.

4.2 Tunnel problem

The tunnel problem is a psychological thought experiment where you travel along a single-lane mountain road in a self-driving car that is fast approaching a narrow tunnel. Just before entering the tunnel a child errantly runs into the road and trips in the center of the lane, effectively blocking the entrance to the tunnel. The car is unable to brake in time to avoid a crash. It has but two options: hit and kill the child, or swerve into the wall on either side of the tunnel, thus killing you [15].

4.3 Standard traffic situation

The traffic situations mentioned in this thesis are situations of unavoidable harm where the following parties are involved:

- A self-driving car with a person behind the wheel that will be referred to as the cars occupant, this car does not have any other passengers.
- One or more traffic participants, for example a person in a car, a pedestrian or a child playing in the streets.

5. Materials and methods

We conducted an online survey constructed with limeSurvey using the Radboud-branded template which can be found in the appendix. For this project we asked around 150 participants to fill out the survey without any compensation. Out of the 150 people 102 started the survey and 40 completed it. At the start of the survey, participants were asked to fill in their age, gender, to what degree they are religious and the biggest influence on their views on self-driving cars, for example the news or work experience. The main purpose of these questions was that previous studies have proven that such factors have a great influence on people's attitude towards self-driving cars [5]. Gender turned out to have the biggest influence, where males had a more positive attitude towards self-driving cars. Younger and less religious people also turned out to have a better attitude towards self-driving cars, but these effects where not as great as gender. In this project, results will be controlled for gender, age, religiosity and view.

5.1 Survey

Participants were presented with 10 scenarios in a randomized order. These scenarios are each part of one of three scenario groups which each have a different setup in order to cover different types of situations that could occur in traffic. These scenario groups are labeled "party A vs party B", "advantage vs risk" and "occupant vs party A". Each scenario in a group differs based on certain factors which pertain to the traffic participants involved in the

scenario and the circumstances and consequences of an accident for each party involved. These factors were selected based on informal inquiry on which aspects could potentially influence scores on fairness. First, a short explanation for each of the factors.

5.2 Factors

Effort for safety
How much effort did a traffic participant put into securing his or her safety? For example, they could spend money on a helmet or a safer car, or the opposite: they could be careless and not wear a seatbelt or not look left and right when crossing a street.

Vulnerability
This is related to the chances and severity of a possible injury to a traffic participant and their ability to possibly prevent injury. For example children and animals are more likely to sustain more serious injuries then adults in similar accidents. Children might also have fewer capabilities to remove themselves from a dangerous situation.

Discrimination
Discrimination in this context would mean that one group of traffic participants would be targeted over another. This could be based on gender, age or more trivial things as wearing a helmet or not.

Additional harm
Besides the direct harm caused by a physical injury to a traffic participant injuries could also lead to additional emotional harm of either the traffic participant themselves or for example family. There could also be additional damage to someone's property or in some cases there could be a more general social impact e.g. a feeling of unsafety as well.

Authority
Authority in this case means accepting traffic regulations set by the authorities, and not changing our behavior if it is suggested that it would reduce the risk in traffic for either themselves or others. For example: the occupant could reduce the risk of an accident if he would drive below the speed limit. But if driving at the speed limit would cause unnecessary risk, the government would most likely intervene by lowering the speed limit. But we assume that the authorities have already made a careful consideration of the advantages and disadvantages of setting a certain speed limit. If this does bring with it a certain amount of risk, authorities have decided that the risk is acceptable. So what motivation could the occupant have to question this decision and take measures of his own to reduce the risk?

Justice
In this case we will define justice as a traffic participant's past in term of moral or immoral behavior. If the choice would be to either collide with a convicted murderer, or with a person who has never committed a criminal offence, it might feel like colliding with the innocent person is unfair because it harms the party that has no history of committing immoral actions.

These factors were randomly assigned over the suitable scenario groups. In each scenario group participants were questioned about:

- The fairness of each algorithm on a scale of -5 to 5: very unfair to very fair.
- How much each relevant factor influenced their scores on fairness.
- If there were any other factors that were not mentioned in the previous questions that influenced their scores.
- How they would divide 100 points between cars that each had one of the algorithms implemented based on how willing they were to pay to use each car.
- When deciding how much they were willing to pay for the use of each car, how their own interest influenced their scores.
- What they considered to be the ideal distribution of collision chance.

The following paragraphs contain a more detailed description for each scenario group.

5.3 Party A vs Party B scenario group

This group consists of 4 scenarios where the participant has to give a rating on fairness for 4 algorithms that each distribute the chance of colliding between a party A and B:

- Collide with party A 100% of the time.
- Collide with party B 100% of the time.
- Collide with party A 50% of the time and collide with party B 50% of the time.
- Collide with party A and B in such a way that the damage over time between the two groups is equal.

In this scenario group the parties will differ on the following factors: age, gender, vehicle safety rating and occupation. The first 3 factors will influence the chance of injury and the last factor will influence the decrease in happiness resulting from a similar injury. The stories for each scenario as used in the survey are mentioned in the following paragraphs.

5.3.1 Gender scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two pedestrians appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will hit a female pedestrian, or swerve to the left, where it will hit a male pedestrian. Both pedestrians are about the same age and it is known that between a similar-aged woman and man, the woman is 25% more likely to be injured. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

5.3.2 Age scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two pedestrians appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will hit a 20 year old male pedestrian, or swerve to the left, where it will hit a 70 year old male pedestrian. It is known that between two traffic participants of these ages, the 70 year old is 200% more likely to be injured. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

5.3.3 Vehicle safety rating scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two vehicles appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will collide with a vehicle with a safety rating of 5 stars*, or swerve to the left, where it will collide with a vehicle with a safety rating of 1 star. Both these vehicles contain only the driver. It is known that between two vehicles with these ratings, the driver of the vehicle with the 1-star rating is 200% more likely to be injured. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

5.3.4 Occupation scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two pedestrians appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will hit a pedestrian who is a professional athlete, or swerve to the left, where it will hit a pedestrian who has an office job. Both pedestrians are about the same age and will sustain similar injuries when hit. After the crash, the athlete's overall happiness will drop 200% more than the office worker because the injuries will get in the way of their job. In both cases, your car will sustain minimal damage and you will not be injured.

In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

5.4 Occupant vs party A scenario group

The second group contains 3 scenarios where the participant has to give a rating on fairness for 4 algorithms. Each algorithm gives the cars occupant a disadvantage in terms of a longer drive in exchange for an advantage in terms of reduced risk for a party A but each to a different degree:

- The car always drives at the speed limit/takes the regular route.
- The car always drives under the speed limit/takes the detour, but there is less chance of colliding with party A.
- The car only drives at the speed limit/takes the normal route when the occupant is late.
- The car only drives under the speed limit/takes the detour under circumstances where the risk for party A is less.

Factors chosen for this scenario group are vulnerability of children and animals and authority. The stories for each scenario as used in the survey are mentioned in the following paragraphs.

5.4.1 Home zone scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. A lot of maintenance is being done on the main roads so the car is forced to drive through a home zone indicated by the following traffic sign:

These areas often house a lot of young families and are designed for children to play and so the speed limit is usually 20 km/h. Because these streets are usually full of parked cars on the sides there is a small chance a small child will suddenly run in front of the self-driving car without being spotted by the systems in time to avoid collision. Colliding will result in moderate injuries to the child and no damage to the vehicle and yourself. It is proven that by driving under the speed limit the chances and the severity of the damage to a child will be less. The downside of that is that the trip would take more time. In the case of the self-driving car driving through living streets, the computer engineers can program the car for four options. Please answer the questions below.

5.4.2 Highway scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. On your daily commute to work the car has to pass a section of highway that has one of the highest rates of fatal car accidents. The car has the option to take a detour in order to avoid this section, but it would take significantly longer. If the car were to get into an accident on the highway not only you, but a lot of other traffic participants would be at risk since the accident will most likely involve multiple vehicles. The chances of an accident for the detour are negligible. In this case, the computer engineers can program the car for four options. Please answer the questions below.

5.4.3 Deer scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. While traveling the car encounters the following traffic sign:

This sign is meant to warn the driver about deer crossing the road. Colliding with a deer will result in the death of the animal and minor damage to the vehicle. It is known that if the car drives slower, the chances of an accident are reduced but as a result of this the trip will take longer. When the car encounters such a sign, the computer engineers can program the car for four options. Please answer the questions below.

5.5 Advantage vs risk scenario group

The third group consists of 3 scenarios where the participant has to give a rating on fairness for 5 algorithms that each distribute the chance of colliding between a party A and a wall which would harm the occupant:

- Collide with party A 100% of the time.
- Collide with party A 75% of the time, collide with a wall 25% of the time.
- Collide with party A 50% of the time, collide with a wall 50% of the time.
- Collide with party A 25% of the time, collide with a wall 75% of the time.
- Collide with a wall 100% of the time.

Factors chosen for this group are effort for safety, discrimination, additional harm, authority, effort for safety and justice. The stories for each scenario as used in the survey are mentioned in the following paragraphs.

5.5.1 Father scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly someone appears in the direct path of the car. The car has the option to: swerve to the right or left, where it will collide with a wall, or to do nothing and collide with this person. The chances of yourself and this person being injured are equal, but this person is known to be a single parent of two young children which will subsequently be harmed emotionally in the event that the person is injured. In preparing the car for such an eventuality, the computer engineers can program the car for five options. Please answer the questions below:

5.5.2 Escaped convict scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly someone appears in the direct path of the car. The car has the option to: swerve to the right or left, where it will collide with a wall, or to do nothing and collide with this person. The chances of yourself and this person being injured are equal, but this person is known to be an escaped convict who was serving a life sentence in prison. In preparing the car for such an eventuality, the computer engineers can program the car for five options. Please answer the questions below:

5.5.3 Inattentive person scenario

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly someone appears in the direct path of the car. The car has the option to: swerve to the right or left, where it will collide with a wall, or to do nothing and collide with this person. The chances of yourself and this person being injured are equal, but this person was only paying attention to the phone and breaking traffic rules by crossing the street where it is not allowed. In preparing the car for such an eventuality, the computer engineers can program the car for five options. Please answer the questions below:

6. Hypotheses and expected patterns

Here we will discuss the hypotheses for each research question.

6.1 Hypotheses on most fair algorithms

Research question: to what extent do people think certain choices made by self-driving cars in ethical dilemmas are fair?

Hypothesis: for the party A vs party B scenario group the equal distribution of damage will be preferred. We think this would be the most likely outcome because other studies show that people have a positive attitude towards utilitarian cars [5], which would make them select the algorithm that always targets the strongest party. On the other hand, this stands in direct opposition of the aversion people have towards algorithms that always target one group over another [3] and equal distribution of damage might be a decent compromise.

If this algorithm is not preferred, this might be because:

- You discriminate against one group because you collide with them more often. This would mean they are more inclined towards the 50/50 algorithm.
- Participants prefer the utilitarian choice, where the least vulnerable party is always hit.
- Participants didn't fully understand the equal damage distribution.

For the advantage vs risk scenario group the algorithm where the car only drives under the speed limit/takes the detour under circumstances where the risk for party A is less will be preferred, because this algorithm splits the disadvantage between the occupant and party A.

If this algorithm is not preferred, this might be because:

- Authorities have deemed the risk that comes with driving at the set speed limit as acceptable, so participants will not be inclined to take measures to lower risk for others themselves.
- Participants value their own interests more than those of others.

For the occupant vs part A scenario group, the hypothesis will differ for each scenario. For the father scenario the 'collide with party A 50% of the time, collide with a wall 50% of the time' algorithm will be preferred, because the collide chance is evenly divided between the father and the occupant

If this algorithm is not preferred, this might be because:

- Participants let the additional emotional harm of the children influence their rating for fairness and so will be more inclined to prefer the 25/75 distribution.
- Participants value their own interests more than others.

For the escaped convict and inattentive person scenario the ‘collide with party A 75% of the time, collide with a wall 25% of the time’ algorithm will be preferred, because participants will value their own interest more than the convicts.

If this algorithm is not preferred, this might be because:

- Participants do not let the circumstances influence their rating for fairness and so will be more inclined to prefer the 50/50 distribution.
- Participants value their own interests a great deal more than others and so would be more inclined towards the 100/0 distribution.

6.2 Hypotheses on importance of factors

Research question: which factors influence people’s view of fairness the most?

For each applicable factor a question is added to the scenario to determine to which degree participants base their fairness ratings on them.

Hypothesis: the factors we measured for will have the following effects on people’s view of fairness:

- Effort for safety: vehicle safety rating (small influence), careless person (small influence)
- Vulnerability: vehicle safety rating (large influence), age (large influence), gender (small influence), home zone (large influence), deer (medium influence)
- Discrimination: First question group (small influence), third question group (small influence)
- Additional harm: occupation (large influence), father (large influence)
- Authority: second question group (small influence), careless person (small influence)
- Justice: escaped convict (medium influence)

This list shows that vulnerability is probably the most important factor as it will have a large influence in most scenarios.

There is a chance that there are more relevant factors that were overlooked. For this purpose we added a question to each scenario asking the participant to state the most important factor if it was not mentioned previously.

6.3 Hypothesis on fairness as a predictor

Research question: is fairness as viewed by the participants a good predictor for willingness to pay for use?

Hypothesis: Fairness will be a good predictor for willingness to pay for use.

If this is not the case, the influence of self-improvement will probably be bigger than we have predicted. To investigate this we have added a question to the relevant scenarios that will determine the importance of own gain when distributing points for willingness to pay for use. Participants could also have different definitions of fairness.

7. Results

The following tables contain the means for the participants' scores on fairness and willingness to pay for use. Fairness was rated on a scale from -5 to 5 and willingness to pay for use on a scale of 0 to 100 where the scores for the 4 or 5 algorithms were required to add up to 100. The bold scores are the highest scores in the row. For Figure 1, the short descriptions for each algorithm were used where party A is the party that is most likely to be injured e.g. the female, and the older person.

scenario:	gender		age		vehicle safety rating		occupation	
	fairness	pay for use	fairness	pay for use	fairness	pay for use	fairness	pay for use
always collide with party A	-3.55	2.25	-2.93	15.5	-3.85	3.38	-4.2	4.13
always collide with party B	-2.58	9.88	-3.13	9.38	**1.38**	**58**	-3.38	12.25
collide with party A/B 50/50	-0.4	35.63	-0.25	35.75	-2.43	10.63	**0.25**	40.63
equal damage distribution	**0.53**	**52.25**	**0.1**	**39.38**	-0.43	28	-0.25	**43**

Figure 1: means of fairness and willingness to pay for the Party A vs Party B group

scenario:	home zone		highway		deer	
	fairness	pay for use	fairness	pay for use	fairness	pay for use
always drive speed limit/take regular route	-1	17.38	-0.63	23.5	0.75	25.25
always drive under speed limit/take detour	**1.98**	**43.5**	0.33	28.38	0.18	19.75
drive speed limit/take regular route when late	-3.13	3.75	-0.68	12	-1.5	7.38
drive speed limit/take regular route when less risk	0.58	35.35	**1.5**	**36.13**	**1.63**	**47.63**

Figure 2: means of fairness and willingness to pay for use scores for the advantage vs risk group

scenario:	parent		convict		inattentive person	
	fairness	pay for use	fairness	pay for use	fairness	pay for use
always collide with party A	-2.48	21.63	-1.23	32.38	-1.08	33.25
collide with party A/wall 75/25	-2.98	7	-1.85	5.88	-2.3	6.13
collide with party A/wall 50/50	-1.88	13.13	-1.98	14.88	-2.75	15.38
collide with party A/wall 25/75	-2.1	10.88	-2.68	4.13	-2.73	4.88
always collide with wall	**0.58**	**47.38**	**-0.53**	**42.75**	**-0.95**	**40.38**

Figure 3: means of fairness and willingness to pay for use scores for the occupant vs party A group

Factor score means for each scenario group on a scale of 0 to 5:

Party A vs party B scenario group

Gender:	vulnerability = 2.58	discrimination = 2.00	
Age:	vulnerability = 1.65	discrimination = 1.88	
Vehicle safety rating:	vulnerability = **3.65**	discrimination = 1.55	effort for safety = 0.85
Occupation:		discrimination = 1.60	additional harm = 1.53

Advantage vs risk scenario group

Home zone:	vulnerability = **4.40**	authority = 2.15
Highway:	vulnerability = 2.68	authority = 1.78
Deer:	vulnerability = 2.73	authority = 2.13

Occupant vs party A scenario group

Parent:	discrimination = 1.5	additional harm = 1.78
Convict:	discrimination = 1.85	justice = 1.2
Inattentive person:	discrimination = 1.98	effort for safety = 2.15

Own interest scores on a scale from 0 to 5:

Advantage vs risk scenario group

Home zone:	2.2
Highway:	2.25
Deer:	2.13

Occupant vs party A scenario group

Parent:	1.78
Convict:	2.35
Inattentive person:	2.1

Ideal chance distributions for Occupant vs party A scenario group:

Parent scenario:	65% to hit wall and 35% to hit parent
Convict scenario:	54% to hit wall and 46% to hit convict
Inattentive person scenario:	58% to hit wall and 42% to hit inattentive person

Results suggest that for the Party A vs party B scenario group, participants prefer equal damage distribution in terms of fairness with the exception of the vehicle safety rating scenario where participants prefer the algorithm that always hits the safer car. For the second scenario group there is a similar pattern: the algorithm were the cars occupant only gains an advantage when there is less risk for the other party is preferred except for the home zone scenario where participants are inclined to protect the children at all cost. For the occupant vs party A group the algorithm that always collides with the wall is preferred. For all of the scenario groups, participants indicated that of all stated factors, vulnerability had the largest influence when determining their scores for fairness with the exception of the age scenario.

For the last two scenario groups participants estimate the degree to which their own interest influenced their scores on willingness to pay for use 2.15 on average on a scale from 0 to 5. They valued their own interest the most in the convict scenario and the least in the parent scenario. When participant were asked to give their preferred chance distributions for the occupant vs party A scenario group, both the convict and the inattentive person scenario are close to a 50/50 distribution while the parent scenario differs with a 65/35 distribution in favor of the parent. A mixed model analysis of the data also showed a significant interaction between scores on fairness and willingness to pay for use ($F = 138.84$, $p < 0.01$).

8. Discussion

For the party A vs party B scenario group the hypothesis that the equal distribution algorithm is preferred appears to be correct except for the fact that for the vehicle safety scenario the algorithm where the least vulnerable vehicle is always collided with is preferred. The same situation occurs with the advantage vs risk group: the hypothesis is correct for the highway and deer scenarios but not for the home zone scenario where the algorithm that always takes the detour in order to protect the children is preferred.

This can be explained by observing the means of the fairness scores for each algorithm in combination with the factor score means. The highest factor scores in the groups are the vulnerability scores for the vehicle safety and home zone scenarios. This indicates that participants only preferred the algorithm that protects the most vulnerable party to the largest extent, when they assigned high scores to the vulnerability factor for that scenario.

For the occupant vs party A scenario group quite unexpected results were found: the

algorithm that always collided with the wall is preferred in all cases. We can also see that the second highest scores for the convict and inattentive person scenario are assigned to the complete opposite algorithm: the one that always collides with the person. The data seems to suggest that this might be because on one hand we have a group of participants that think that the first priority should be the protection of the cars occupant and on the other hand we have participants that would prefer to collide with the wall because of the emotional damage that they would suffer if they were to collide with the person. But in the additional data there are very few participants who have stated they prefer one algorithm over the other for these reasons, so more research would need to be done on the exact cause. We also asked participant which chance distribution they preferred but the average of all scores is close to a 50/50 division except for the parent scenario. The result for the parent scenario could be explained by the fact that the data suggests that participant prioritize the protection of children, which might suffer from emotional damage in this scenario, shifting the preferred collision chance more towards the wall. The 50/50 pattern is most likely explained by the fairness scores which are pretty evenly split between favoring one side or the other.

For the factors that underline participant's fairness scores, we asked participants if there was a factor they based their scores on but were not included in the survey. From the data it appears that a large portion of participants didn't understand this question. Answers included comments on what participants would like to know about the situation or that a car would never be able to measure the information needed anyway, so very few usable responses were found. The following factors were found: life expectancy of all parties, social impact of the injury to a party, and for the occupant vs party A scenario group the guilt or emotional damage for the cars occupant if it would collide with a person. However, these factors were only mentioned once or twice over 40 participants which raises the question if these factors are relevant for the general population or not. What was also interesting about this data is that some participants stated that 'you should not play god' or 'it is unfair to choose at all', for the first and last question group. This is a bit unexpected since these scenarios also include a 50/50 option for the algorithm. For any future research, this option might be better labeled as random.

As the data showed that there were few factors that were unaccounted for, the factors that were mentioned in the survey should explain the largest portion of the scores on fairness. When we look at the data we can see that vulnerability is the highest scoring factor on all scenarios except for age. So we can conclude that vulnerability is indeed the most important factor.

The data lastly shows that participants fairness scores are indeed a good predictor for scores on willingness to pay for use ($F = 138.84$, $p < 0.01$).

9. Recommendations

Since fairness is indeed a good predictor for willingness to pay for use the recommendations are as follows:

In situations of unavoidable harm where the driverless car would have to collide with either party A or party B, the car should first attempt to make an accurate prediction on the chance and severity of the injury for each party. Then the chance to collide with each party should be determined based on which chances will lead to an even distribution of damage over time.

In situations where the driverless car has the option to reduce risk for a party A in exchange for a disadvantage for the occupant by performing action B as in the second scenario group, the algorithm should check if the risk for party A is reduced by circumstances at the current time and perform action B if this is not the case.

In situations where the driverless car has the option to either collide with a wall or collide with a party A as in the third scenario group, the car should always collide with the wall.

The cars algorithms should also attempt to avoid any type of harm to children. Of course, more research should be done in order to further improve these recommendations.

10. Improvements and future research

The first thing to address for future research would be the participants. When looking at the data participants aged 40-65 take up 75% of all participants. In order to generalize results it might be better to have a more diverse group of participants. The data also had some inconsistent answers in it, for example people with the highest score on the importance of not discriminating in some cases had fairly high scores for fairness for the algorithms that only target one party. This indicates that, for future research, survey data might be improved by recruiting participants with some knowledge of the topic of self-driving cars and moral decisions and are motivated to provide good and consistent answers. It is also noticeable that less than half of the participants who started the survey finished it and some open questions where simply filled out wrong. This indicates that the formulation of questions and the information provided could be improved for future surveys.

Participants seemed reluctant to choose one group over another based on the information given, especially for age and gender. So for a follow up survey a scenario should be added that does not give any information except for the likelihood that each person gets injured. To see if participants would still consider preferring one group over another would be unfair.

There should also be some prior research on the best way to formulate the scenarios as research indicate that wording and framing effects influence moral intuitions [10]. This study shows that people agree more strongly to questions when they were worded to Save than when they were worded to Kill. For this questionnaire this means that fairness scores would most likely be higher when the questions are changed from Kill wording such as “always

collide with party A" to the Save wording "never collide with party B".

Context should also be more precise for example it should be made clear that the severity of all injuries is the same, since some participants were uncertain about his. It is possible that the importance of vulnerability is slightly overestimated because of this.

It is also not completely clear how the underlying factors predict participants' scores on fairness since we only asked the participants how many certain factors influenced their scores on fairness. In order to get an accurate view on how the factors predict fairness scores we would need to measure the perceived vulnerability of each party involved and compare these scores with the corresponding scores on fairness.

10.1 Ethical issues surrounding implementation

When the ideal algorithms in terms of fairness have finally been determined, we should start considering the consequences of these implementations. Let's say vulnerability is the most important factor and that justice proves to be a viable predictor as well as long as the occupant is not at risk. Data needed for this suggested algorithm could include gathering information about your eating habits, how much you exercise, smoke and drink, family medical history, history of traffic violations, age, height, weight, gender, race, occupation and the list goes on. We are still far away from having the option of collecting this data at the speed that is necessary to make the split second decision between possibly harming one party or the other, but car sensors or chips with information about its occupants could go a long way. If we could measure these aspects, would the traffic participants be willing to give up some of their privacy to ensure these algorithms can work properly? And if they don't, are they willing to compromise in terms of how well the algorithms operate? Maybe they would even prefer different algorithms that do not make choices in these kinds of situation in order to not have to compromise their privacy at all. These are all questions that will need to be answered to ensure the acceptance of self-driving cars.

References

[1] Goodall, N. J. (2014). Machine ethics and automated vehicles. In Road Vehicle Automation (pp. 93-102). Springer International Publishing.

[2] Evans, L. (2008). Death in Traffic: Why Are the Ethical Issues Ignored?. Studies in ethics, law, and technology, 2(1).

[3] Goodall, N. (2014). Ethical Decision Making During Automated Vehicle Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2424*, 58–65. http://doi.org/10.3141/2424-07

[4] Lin, P. (2016). Why ethics matters for autonomous cars. In Autonomous Driving (pp. 69-85). Springer Berlin Heidelberg.

[5] Bonnefon, J. F., Shariff, A., & Rahwan, I. (2015). Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?. arXiv preprint arXiv:1510.03346.

[6] Millar, J. (2014, May). Technology as moral proxy: Autonomy and paternalism by design. In Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium on (pp. 1-7). IEEE.

[7] Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 28(4), 15.

[8] Anderson, M., & Anderson, S. L. (Eds.). (2011). Machine ethics. Cambridge University Press.

[9] Millar, J. (2015). Technological Moral Proxies and the Ethical Limits of Automating Decision-Making In Robotics and Artificial Intelligence.

[10] Crisp, R. (2015). The Tunnel Problem. Practical Ethics. (Jul 22). Online:

http://blog.practicalethics.ox.ac.uk/2015/07/the-tunnel-problem/.

[11] Lin, P. (2013). The Ethics of Saving Lives With Autonomous Cars Are Far Murkier

Than You Think." Wired. (July 30). Online: http://www.wired.com/opinion/2013/07/the-surprising-ethics-of-robot-cars/.

[12] Millar, J. (2004) An ethical dilemma: When robot cars must kill, who should pick the victim? http://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/

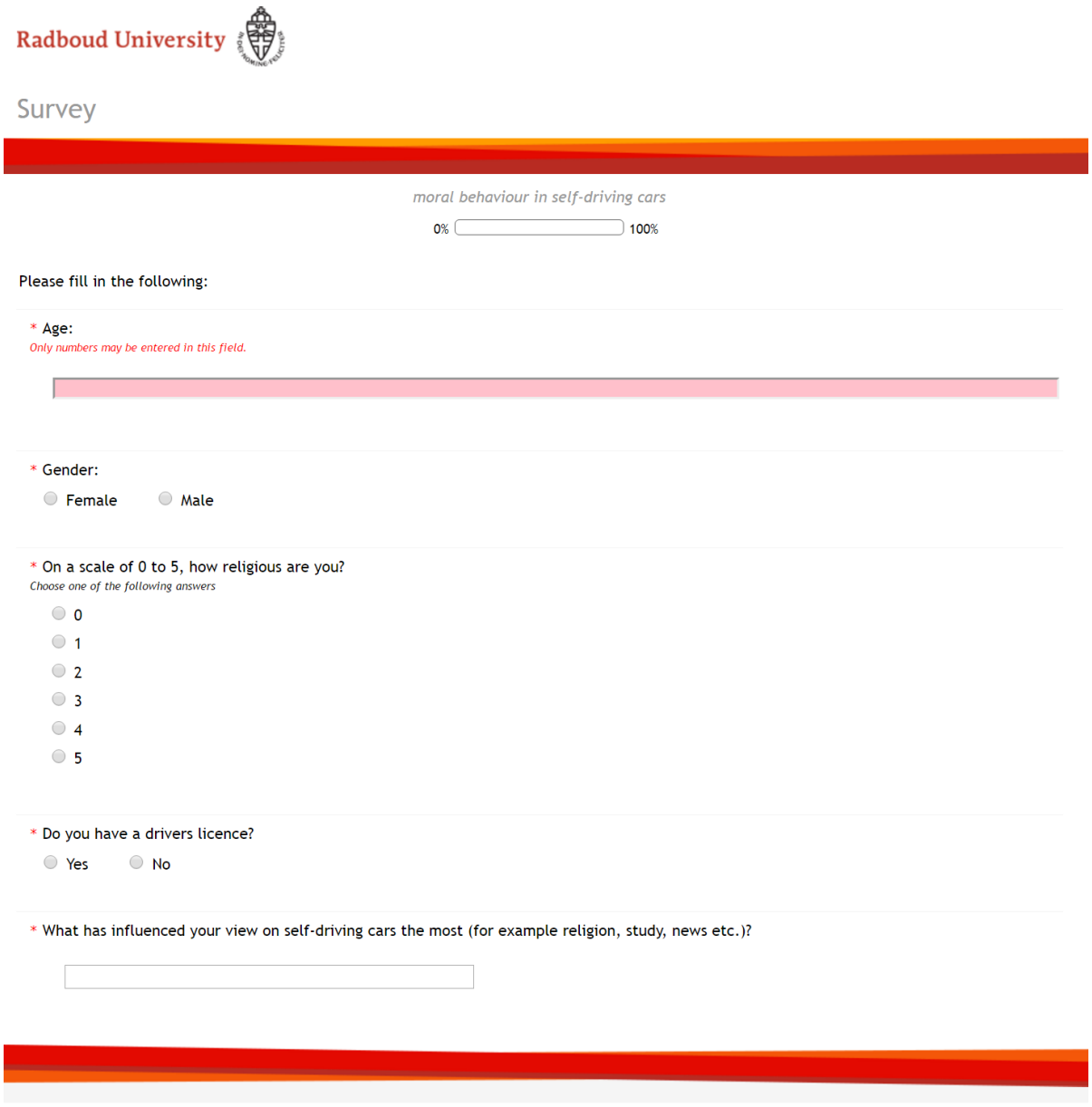
[13] Howard, D., & Dai, D. (2014). Public perceptions of self-driving cars: The case of Berkeley, California. In Transportation Research Board 93rd Annual Meeting (No. 14-4502).

[14] Foot, P. (1967). The problem of abortion and the doctrine of double effect.

[15] Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. Ethology and Sociobiology, 17(3), 145-171.

[16] Millar, J. (2014, May). Technology as moral proxy: Autonomy and paternalism by design. In Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium on (pp. 1-7). IEEE.

Appendix



Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

Please fill in the following:

* Age:

Only numbers may be entered in this field.

* Gender:

Female Male

* On a scale of 0 to 5, how religious are you?

Choose one of the following answers

0
1
2
3
4
5

* Do you have a drivers licence?

Yes No

* What has influenced your view on self-driving cars the most (for example religion, study, news etc.)?

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two pedestrians appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will hit a female pedestrian, or swerve to the left, where it will hit a male pedestrian. Both pedestrians are about the same age and it is known that between a similar-aged woman and man, the woman is 25% more likely to be injured. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Collide with the female pedestrian 50% of the time and collide with the male pedestrian 50% of the time.	○	○	○	○	○	○	○	○	○	○	○
Distribute the chance to collide with the male and female in such a way that the damage done to each group over time will be equal.*	○	○	○	○	○	○	○	○	○	○	○
Always collide with the female pedestrian.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the male pedestrian.	○	○	○	○	○	○	○	○	○	○	○

? **Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One party is more likely to get injured.	○	○	○	○	○	○
One party will be collided with more often than the other.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Collide with the female/male 50/50. 0 100

Always collide with the male. 0 100

Always collide with the female. 0 100

Even damage distribution female/male.* 0 100

? **Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two pedestrians appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will hit a 20 year old male pedestrian, or swerve to the left, where it will hit a 70 year old male pedestrian. It is known that between two people of these ages, the 70 year old is 200% more likely to be injured. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Always collide with the 20 year old pedestrian.	○	○	○	○	○	○	○	○	○	○	○
Distribute the chance to collide with the 70 year old and 20 year old in such a way that the damage done to each group over time will be equal.*	○	○	○	○	○	○	○	○	○	○	○
Collide with the 70 year old pedestrian 50% of the time and collide with the 20 year old pedestrian 50% of the time.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the 70 year old pedestrian.	○	○	○	○	○	○	○	○	○	○	○

? **Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One party will be collided with more often than the other.	○	○	○	○	○	○
One party is more likely to get injured.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Always collide with the 70 year old. 0 100

Collide with the 20/70 year old 50/50. 0 100

Even damage distribution between 20/70 year old.* 0 100

Always collide with the 20 year old. 0 100

? **Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two vehicles appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will collide with a vehicle with a safety rating of 5 stars*, or swerve to the left, where it will collide with a vehicle with a safety rating of 1 star. Both these vehicles contain only the driver. It is known that between two vehicles with these ratings, the driver of the vehicle with the 1-star rating is 200% more likely to be injured. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

* A vehicles safety rating is given on a scale from 1 to 5 stars, with 1 star being very poor and 5 stars being excellent.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Always collide with the vehicle with the 5-star rating.	○	○	○	○	○	○	○	○	○	○	○
Collide with the vehicle with the 1-star rating 50% of the time and collide with the vehicle with the 5-star rating 50% of the time.	○	○	○	○	○	○	○	○	○	○	○
Distribute the chance to collide with the 1-star and 5-star vehicle in such a way that the damage done to each group over time will be equal.*	○	○	○	○	○	○	○	○	○	○	○
Always collide with the vehicle with the 1-star rating.	○	○	○	○	○	○	○	○	○	○	○

? *Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One person paid more for his safety.	○	○	○	○	○	○
One party will be collided with more often than the other.	○	○	○	○	○	○
One party is more likely to get injured.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.

Please fill in at least one answer

Collide with 1-star/5-star 50/50. 0 100

Even damage distribution 1-star/5-star.* 0 100

Always collide with 5-star. 0 100

Always collide with 1-star. 0 100

? *Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly, two pedestrians appear ahead, in the direct path of the car. The car has the option to: swerve to the right, where it will hit a pedestrian who is a professional athlete, or swerve to the left, where it will hit a pedestrian who has an office job. Both pedestrians are about the same age and will sustain similar injuries when hit. After the crash, the athlete's overall happiness will drop 200% more than the office worker because the injuries will get in the way of their job. In both cases, your car will sustain minimal damage and you will not be injured. In preparing the car for such an eventuality, the computer engineers can program the car for four options. Please answer the questions below.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Distribute the chance to collide with the athlete and office worker in such a way that the damage done to each group over time will be equal.*	○	○	○	○	○	○	○	○	○	○	○
Collide with the athlete 50% of the time and collide with the office worker 50% of the time.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the athlete.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the office worker.	○	○	○	○	○	○	○	○	○	○	○

? **Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One person will be emotionally harmed in addition to the physical harm.	○	○	○	○	○	○
One party will be collided with more often than the other.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.

Please fill in at least one answer

Even damage distribution athlete/office worker.* 0 100

Always collide with the office worker. 0 100

Collide with the athlete/office worker 50/50. 0 100

Always collide with the athlete. 0 100

? **Equal damage distribution. For example: if party A is twice as likely to get injured, they should be hit half as often to compensate for this.*

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. A lot of maintenance is being done on the main roads so the car is forced to drive through a home zone indicated by the following traffic sign:

These areas often house a lot of young families and are designed for children to play and so the speed limit is usually 20 km/h. Because these streets are usually full of parked cars on the sides there is a small chance a small child will suddenly run in front of the self-driving car without being spotted by the systems in time to avoid collision. Colliding will result in moderate injuries to the child and no damage to the vehicle and yourself. It is proven that by driving under the speed limit the chances and the severity of the damage to a child will be less. The downside of that is that the trip would take more time. In the case of the self-driving car driving through living streets, the computer engineers can program the car for four options. Please rate each option on morality, fairness and willingness to pay for use.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
The car always drives at the speed limit.	○	○	○	○	○	○	○	○	○	○	○
The car always drives under the speed limit.	○	○	○	○	○	○	○	○	○	○	○
The car only drives at the speed limit on times that less children are known to play outside.	○	○	○	○	○	○	○	○	○	○	○
The car only drives at the speed limit when you are running late.	○	○	○	○	○	○	○	○	○	○	○

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
The authorities set these traffic rules and there should be no reason to question them.	○	○	○	○	○	○
Children are vulnerable and should be protected.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Always drive at the speed limit. 0 100

Always drive under the speed limit. 0 100

Drive at the speed limit when you're late. 0 100

Drive at the speed limit when there's less risk. 0 100

* To what degree did your own interest influence your scores for willingness to pay for use?
Choose one of the following answers

- ○ 0: no influence
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5: large influence

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. On your daily commute to work the car has to pass a section of highway that has one of the highest rates of fatal car accidents. The car has the option to take a detour in order to avoid this section, but it would take significantly longer. If the car were to get into an accident on the highway not only you, but a lot of other traffic participants would be at risk since the accident will most likely involve multiple vehicles. The chances of an accident for the detour are negligible. In this case, the computer engineers can program the car for four options. Please rate each option on morality, fairness and willingness to pay for use.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
The car always takes the highway.	○	○	○	○	○	○	○	○	○	○	○
The car always takes the detour.	○	○	○	○	○	○	○	○	○	○	○
The car only takes the highway when you are running late.	○	○	○	○	○	○	○	○	○	○	○
The car only takes the highway when it has less traffic.	○	○	○	○	○	○	○	○	○	○	○

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
The authorities set these traffic rules and there should be no reason to question them.	○	○	○	○	○	○
There is a risk for a large group of people and this should be avoided.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
The authorities set these traffic rules and there should be no reason to question them.	○	○	○	○	○	○
There is a risk for a large group of people and this should be avoided.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Always take the detour.
0 100

Always take the highway
0 100

Only take the highway when you're late.
0 100

Only take the highway when it has less traffic.
0 100

* To what degree did your own interest influence your scores for willingness to pay for use?
Choose one of the following answers

- ○ 0: no influence
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5: large influence

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. While traveling the car encounters the following traffic sign:

This sign is meant to warn the driver about deer crossing the road. Colliding with a deer will result in the death of the animal and minor damage to the vehicle. It is known that if the car drives slower, the chances of an accident are reduced but as a result of this the trip will take longer. When the car encounters such a sign, the computer engineers can program the car for four options. Please rate each option on morality, fairness and willingness to pay for use.

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
The car always drives at the speed limit.	○	○	○	○	○	○	○	○	○	○	○
The car always drives under the speed limit.	○	○	○	○	○	○	○	○	○	○	○
The car only drives at the speed limit when you are running late.	○	○	○	○	○	○	○	○	○	○	○
The car only drives at the speed limit at times when deer are less likely to cross.	○	○	○	○	○	○	○	○	○	○	○

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
The authorities set these traffic rules and there should be no reason to question them.	○	○	○	○	○	○
Deer are vulnerable and should be protected.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are four identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 4 scores add up to 100 (for example (25, 25, 25, 25), (0,100,0,0) or (0,60,30,10)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Always drive under the speed limit. 0 100

Drive at the speed limit when there's less risk. 0 100

Drive at the speed limit when you're late. 0 100

Always drive at the speed limit. 0 100

* To what degree did your own interest influence your scores for willingness to pay for use?
Choose one of the following answers

- ○ 0: no influence
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5: large influence

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly someone appears in the direct path of the car. The car has the option to: swerve to the right or left, where it will collide with a wall, or to do nothing and collide with this person. The chances of yourself and this person being injured are equal, but this person is known to be a single parent of two young children which will subsequently be harmed emotionally in the event that the person is injured. In preparing the car for such an eventuality, the computer engineers can program the car for five options. Please answer the questions below:

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Collide with the person 50% of the time, collide with a wall 50% of the time.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the wall.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the person.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 25% of the time, collide with a wall 75% of the time.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 75% of the time, collide with a wall 25% of the time.	○	○	○	○	○	○	○	○	○	○	○

If you could divide 100% between both parties, what would you think is the ideal distribution for the collision chance? Please make sure the 2 scores add up to 100%, for example: (100,0), (60,40) or (95,5).

Please click and drag the slider handles to enter your answer.

Please fill in at least one answer

Collision chance person:

Collision chance wall:

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One party will be collided with more often than the other.	○	○	○	○	○	○
The children will be emotionally harmed in addition to the physical harm of the person.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are five identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these five options based on how likely you would be to pay to use each one. Please make sure that the 5 scores add up to 100 (for example (20, 20, 20, 20,20), (0,100,0,0,0) or (0,60,30,10,0)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Always collide with the wall.
0 100

Collide with the person/the wall 25/75.
0 100

Collide with the person/the wall 75/25.
0 100

Collide with the person/the wall 50/50.
0 100

Always collide with the person.
0 100

* To what degree did your own interest influence your scores for willingness to pay for use?
Choose one of the following answers

- ○ 0: no influence
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5: large influence

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly someone appears in the direct path of the car. The car has the option to: swerve to the right or left, where it will collide with a wall, or to do nothing and collide with this person. The chances of yourself and this person being injured are equal, but this person is known to be an escaped convict who was serving a life sentence in prison. In preparing the car for such an eventuality, the computer engineers can program the car for five options. Please answer the questions below:

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Always collide with the person.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 75% of the time, collide with a wall 25% of the time.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 50% of the time, collide with a wall 50% of the time.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 25% of the time, collide with a wall 75% of the time.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the wall.	○	○	○	○	○	○	○	○	○	○	○

If you could divide 100% between both parties, what would you think is the ideal distribution for the collision chance? Please make sure the 2 scores add up to 100%, for example: (100,0), (60,40) or (95,5).

Please click and drag the slider handles to enter your answer.

Please fill in at least one answer

Collision chance person:

Collision chance wall:

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One party will be collided with more often than the other.	○	○	○	○	○	○
The person has done 'bad' things in the past .	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are five identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these five options based on how likely you would be to pay to use each one. Please make sure that the 5 scores add up to 100 (for example (20, 20, 20, 20,20), (0,100,0,0,0) or (0,60,30,10,0)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Collide with the person/the wall 50/50. 0 100

Always collide with the wall. 0 100

Collide with the person/the wall 75/25. 0 100

Collide with the person/the wall 25/75. 0 100

Always collide with the person. 0 100

* To what degree did your own interest influence your scores for willingness to pay for use?
Choose one of the following answers

- ○ 0: no influence
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5: large influence

Radboud University

Survey

moral behaviour in self-driving cars

0% 100%

You are the sole passenger in an autonomous self-driving vehicle traveling at the speed limit. Suddenly someone appears in the direct path of the car. The car has the option to: swerve to the right or left, where it will collide with a wall, or to do nothing and collide with this person. The chances of yourself and this person being injured are equal, but this person was only paying attention to the phone and breaking traffic rules by crossing the street where it is not allowed. In preparing the car for such an eventuality, the computer engineers can program the car for five options. Please answer the questions below:

* On a scale from -5 to 5 how fair are the following options?

	very unfair			somewhat unfair		neutral		somewhat fair			very fair
	-5	-4	-3	-2	-1	0	1	2	3	4	5
Collide with the person 25% of the time, collide with a wall 75% of the time.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the wall.	○	○	○	○	○	○	○	○	○	○	○
Always collide with the person.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 75% of the time, collide with a wall 25% of the time.	○	○	○	○	○	○	○	○	○	○	○
Collide with the person 50% of the time, collide with a wall 50% of the time.	○	○	○	○	○	○	○	○	○	○	○

If you could divide 100% between both parties, what would you think is the ideal distribution for the collision chance? Please make sure the 2 scores add up to 100%, for example: (100,0), (60,40) or (95,5).

Please click and drag the slider handles to enter your answer.

Please fill in at least one answer

Collision chance person:

Collision chance wall:

* On a scale from 0 through 5 how much did the following factors influence your scores on fairness from the previous question?

	0	1	2	3	4	5
One party will be collided with more often than the other.	○	○	○	○	○	○
The person is the cause of the dangerous situation.	○	○	○	○	○	○

* If not one of the factors above, what was the most important consideration that you based your scores on fairness on?

Assume that there are five identical cars that all cost the same, but each car has one of these algorithms implemented. Please divide 100 points between these four options based on how likely you would be to pay to use each one. Please make sure that the 5 scores add up to 100 (for example (20, 20, 20, 20,20), (0,100,0,0,0) or (0,60,30,10,0)).

Please click and drag the slider handles to enter your answer.
Please fill in at least one answer

Collide with the person/the wall 50/50. 0 100

Collide with the person/the wall 75/25. 0 100

Always collide with the wall. 0 100

Collide with the person/the wall 25/75. 0 100

Always collide with the person. 0 100

* To what degree did your own interest influence your scores for willingness to pay for use?
Choose one of the following answers

- ○ 0: no influence
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5: large influence