A Study by Simulation on the Moral Dilemmas surrounding Autonomous Cars

Michelle Falzon
University of Malta
michelle.falzon.14@um.edu.mt

Julian Abela
University of Malta
julian.abela.14@um.edu.mt

Keith Camilleri
University of Malta
keith.camilleri.14@um.edu.mt

Kristina Catania
University of Malta
kristina.catania.14@um.edu.mt

Jenny Attard
University of Malta
jenny.n.attard.14@um.edu.mt

Abstract

The aim of this study is to determine how autonomous vehicles should be programmed to react in the event of a car accident where harm to another life is unavoidable. With the use of a web-based simulation, participants were placed in the driver's seat of a car and shown variations of the trolley problem. They were allowed to choose the accident's outcome in a short amount of time. The results concluded that there were a good number of people who would choose to sacrifice themselves over the lives of others. When it comes to decisions between lives of pedestrians, animals and old people were least likely to be saved, while children and pregnant women were most likely to be saved. It was also concluded that the utilitarian approach is very prevalent in this type of scenario, with the vast majority of participants choosing to save the highest number of lives.

1. Introduction

Recent advancements in technology have led to the invention of self-driving (or autonomous) cars. The motivation behind these vehicles is to reduce traffic, pollution and road accidents. However, an important ethical issue surrounds this new technology; if the car comes across a case in which harm is unavoidable, should it save its passengers and make the decision to injure others, or should it sacrifice the passengers and avoid injuring other drivers or pedestrians? Given that there's no option to avoid any injury in such situations, who is responsible for the injuries/deaths that occur? Would people be inclined to buy such a car?

These questions all stem from how the underlying algorithm of the self-driving car works. Since machines make calculations faster than humans, they lack a sense of instinct, and will have to be programmed to make such decisions consciously based on logic. Bonnefon et al. [Bonnefon et al., 2016], have already attempted to solve this issue of morality through various online surveys explained in more detail in section 2.

In this study, we aim to build upon their work by introducing new cases including differently aged people and other cars as well as providing the option to sacrifice yourself instead of others. This will be implemented by creating a game in which the user must decide who/what to hit in each scenario. This is preferred over a survey as it will lead to more reliable results by giving the players the opportunity to see the people being hit, and thus being able to easily picture themselves in the situation and react as they would in real life, rather than simply reading text. The data collected from volunteers will be used to study how people react in cases of unavoidable harm, to determine how these self-driving cars could be programmed.

2. Scientific Review

In their paper [Bonnefon et al., 2016], J. Bonnefon, A. Shariff and I. Rahwan aimed to study people's judgements in order to reach conclusions on the appropriateness of programming autonomous vehicles (AVs) to follow a utilitarian approach (maximizing the number of lives saved). They conducted numerous surveys, each aimed at collecting data on the moral compass of participants.

In their first study, the participants were asked if driverless cars should sacrifice the driver in order to prevent the deaths of 10 pedestrians in circumstances where harm is inevitable. Their work showed that the majority (76%) agreed that it is morally acceptable for an autonomous vehicle (AV) to adopt such an approach.

In the second experiment, the same situation was presented but the number of pedestrians varied. The results showed that they did not agree on sacrificing the driver if only one pedestrian's life was at stake. This attitude changed when more pedestrian's lives were involved.

Surprisingly, their opinions did not shift when they were asked to imagine that a family member or friend was in the car with them.

In the next study, participants were asked to allocate points to different algorithms depending on the morality of the algorithm, how comfortable they were with having AVs programmed that way and how likely it is that they would buy such a vehicle. The algorithm which swerved into a single person to save 10 people was preferred, while the algorithm which swerved into a pedestrian to save another pedestrian received few points. The algorithm which killed its passenger to save 10 lives received a mixture of opinions. The results showed that people thought utilitarian cars are ideal for others to buy, but would not invest in one themselves.

In their previous paper, [Bonnefon et al., 2015] they argue that an adoption to self-driving cars guarantees that traffic accidents will decrease significantly, since most of these can be attributed to human error. An example of cars which have succeeded in performing very well in real environments are those made by Google, now called Waymo [Waymo], which travelled millions of miles around the United States.

However, it is very difficult to avoid all accidents that can occur in the complexities of the real world, and some decisions need to be taken by the car itself, such as, the most ideal direction to crash into. Manufactures and regulators should aim to achieve 3 goals; firstly that of being consistent, secondly keeping public outrage to a minimum, and finally encouraging buyers.

As already mentioned, some accidents are unavoidable and it is in those cases that potential buyers are mostly concerned about. When buying a car, one of the features that a buyer looks for is safety. However, if the car's system is utilitarian, people may be discouraged from buying because the idea makes them uneasy.

Therefore, manufacturers and regulators should study the results of surveys that employ this kind of experiment. Multiple surveys were conducted to study users' decisions in unavoidable accidents. Participants were asked to input basic information such as age, sex and religion.

In the first study, around 400 participants were asked to take part in a survey. They were given a scenario where autonomous car, which does not have time to stop, comes across pedestrians randomly crossing the street. They were asked if the car should swerve into a barrier, and therefore, kill its passenger or keep going straight and hit pedestrians (killing either 1 or 10 pedestrians). It was concluded that the most preferred action was to maximize the number of lives saved, even when it meant killing the passenger. They were also asked if they would consider that, in such scenarios, the driver must always ensure that the most lives are saved even if s/he must die should be enforced by law. Participants were more willing to consider legal enforcement of the sacrifice when the decision was made by a self-driving car, rather than by a human driver.

In study 2, the researchers wished to determine if people would consider buying a car which always aimed to save the most lives, even if the driver must die. Although people said that they would be supportive for others to buy such cars, they would not buy one for themselves. The results also showed that males were more likely to be interested in buying such a vehicle.

In the final study, participants were instructed to imagine themselves in the car, or to imagine an anonymous character. They were once again asked if the car should swerve or remain on its course as well as how they expected future driverless cars to be programmed. Three fourths of the participants believed that autonomous vehicles should self-sacrifice to save more lives. However, a smaller percentage believed that self-driving cars will be programmed that way by manufacturers.

Such studies show that people react to situations in different ways and thus there is no clear answer as to how these vehicles should be programmed. However, due to the increasing amount of traffic accidents, the importance surrounding autonomous vehicles is growing, with hopes that they will reduce the frequency of the accidents as well as the damage done. Thus, it is important to keep searching to find a solution to the reasoning these vehicles should adopt in cases of unavoidable harm.

One drawback from the previously mentioned papers is that participants were never made fully aware of the rush of emotions experienced during a situation such as the Trolly Dilemma. These emotions might interfere with the thought process and might cause people act differently than when presented with the same situation on paper.

C.D Navarrete et al, [Navarrete et al., 2012] studied this phenomena by setting up a Virtual Reality (VR) environment where people were able to experience this dilemma through their own eyes without causing harm to anyone.

In the scene depicted through the VR, participants were shown a boxcar track. The track on which the boxcar was driven had five workmen on it, while another track had only one man. The participants had to choose the fate of those men by selecting a rail for the boxcar to drive on.

The experiment was posed in one of two different ways for each participant. The 'action condition' experiment tested to see whether participants would pull the lever to move the boxcar from the rail of five to the rail of one, i.e. taking the utilitarian approach. In the 'omission condition' experiment, the roles were reversed. If the participants refused to take any action, the boxcar kept moving straight and killed one man. The results showed that the Utilitarian approach dominated whether this was done actively (by pulling the lever) or passively. In the experiments, 90.5% and 88.5% of the participants respectively chose to save more lives.

Alexander Skulmowski et al. [Skulmowski et al., 2012] created a virtual reality experiment to investigate the decisions in the context of sacrificing a victim in a trolley dilemma scenario. Different characters were used in the scenario to represent the different gender, ethnicity and orientation in space.

This paper mentions that a recent study conducted by Friedman et al. (2014) shows that in order to investigate any changes in the decision-making process, repeated presentations of the same scenarios are shown. Building on what Friedman et al. found, this study investigates the effects people have when they are faced with the same scenario multiple times, to analyse if there are any changes in the decisions made. The first few confrontations with the dilemma leads to rather spontaneous and automatic responses. The repetition gave participants time to reflect upon their decision tendencies in a controlled manner.

Eye tracking was used to analyse participants' gaze behaviours whilst the use of virtual reality (VR) equipment helped in capturing the precise time when the decision was taken. By using VR, emotions were also recorded. The paper further informs the reader that Seidel and Prinz (2013) showed that negative emotions are induced by playing harsh noises, which affects the moral judgment. In addition, this anger and negativity lead to a higher possibility of people opting to kill more avatars as a sense of relief, in order to cope with this emotion.

A total of 66 students took part in the study. Their age ranged from 18 to 30 years old with an average age of 22. Participants were faced with 4 dilemmas in which different gender, ethnicity, body orientation, and the number of avatars present in the scenario were varied. When a faced with making the choice to kill a group or just one person, almost all the participants saved the group of people by sacrificing the single person. This was true for both genders taking part in this experiment.

Both males and females showed a preference towards killing less females. Additionally, the study concluded that when participants were faced with a group of avatars in the dilemma, the response was faster, which suggests that utilitarian decisions can be made in a shorter amount of time.

Li et al. [Li et al, 2016] tackle two main issues, the first of which being "To which target will people allocate responsibility for an accident involving an autonomous car - to the producer, the owner, or the car itself?", while the other being "Do people expect autonomous vehicles to take utilitarian actions when facing moral dilemmas?". In both situations, the participants were given one from six scenarios to read and answer.

The first study was closely linked to the responsibility of actions. Participants were given scenarios where very similar situations were considered, for example having one case where the car was fully autonomous while in the other the car was driven by a human. Results from these statistics showed that participants held the human driver much more responsible when compared to the autonomous vehicle, even when the vehicle itself was at fault. In the case where the fault was of the pedestrian, both the autonomous car and the human-driven car scored low responsibilities. Participants felt that the government should be held responsible in the case of autonomous vehicles. This left participants with an uneasy feeling at the thought of being able to purchase such a vehicle.

The second study targeted whether or not such autonomous vehicles should always consider the utilitarian approach. Three car situations were considered; the autonomous car, the normal human-driven car, and a handover car where the human driver takes control of the vehicles during the time of action. Similar to the first study, participants were given text descriptions which described the trolley problem. Results from this study showed that three thirds of the participants opted for the utilitarian approach. Evaluating results closer showed that 82.5% in the autonomous car would have opted for the utilitarian approach, 75% in the handover, and 67.5% in the human-driven car.

In conclusion to both studies above, the autonomous vehicles never seemed to be held responsible for their actions, but rather the manufactures who programmed such vehicles. Moreover, participants also opted for the utilitarian approach in order to save five lives and sacrifice the one.

Malle et all. [Malle et al, 2015] study and compare the morality of the choices taken by human and robotic agents facing the same moral dilemma. They aim to compare the levels of blame, permissibility and wrongness, given by participants, of the actions taken.

Given tough choices, people tend to weigh in norms, emotions and consequences to determine the best possible action to take. However, they might also reach the decision unconsciously, based on intuition which cannot be fully justified. But how should robotic agents tackle such decisions?

Typical studies aim to reveal the norms applicable to a situation by asking volunteers to indicate whether an action is acceptable, permissible or one they would have chosen. In their experiments, Malle. et al., extended this by showing the participants the same situation with a robot (rather than a human) performing the action. This was done in order to test whether the same norms would apply to robot agents.

Two experiments were carried out. In the first, participants were shown a scenario involving 4 miners working on a track in a coal mine, with a train heading straight towards them. They were asked to take the bystander approach by asking if it was permissible for the repair man (or robot) to redirect the train to another track and kill a single miner. They were also asked to explain the reasoning behind their choice.

The same scenario was presented to them again. This time, they were told which action the repair man (or robot) had chosen. They were asked to specify whether the agent should be blamed for taking such an action, as well as the reasoning behind their decision. The same scenarios were presented to the participants using the other agent instead.

The results showed that both humans and robots were deserving of more blame when they took action to direct the train and sacrifice one person. However, the actions of robotic agents were deserving of more blame more for not taking action when compared to humans.

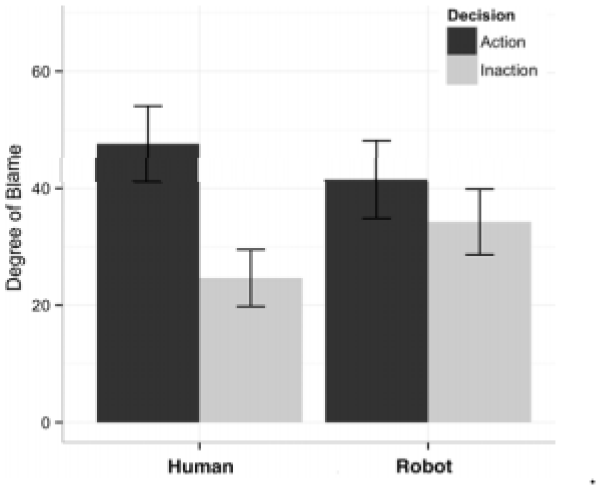

Figure 1. Rates of blame in Experiment 1 as a function of agent type and the agent's decision (to divert the train or not)

In the second experiment, the same scenario was provided to the participants. The action taken was specified, and the participants had to determine whether the action was morally wrong or not, as well as the reason behind their choice. They then were also asked to answer these questions for the same scenario using the other agent. In both cases the same action was chosen (ex: divert train).

The results showed that less people considered the sacrifice taken by the robot to save 4 people as morally wrong when compared to the human that made the same choice. In contrast, the human agent was blamed less for inaction than the robotic agent.

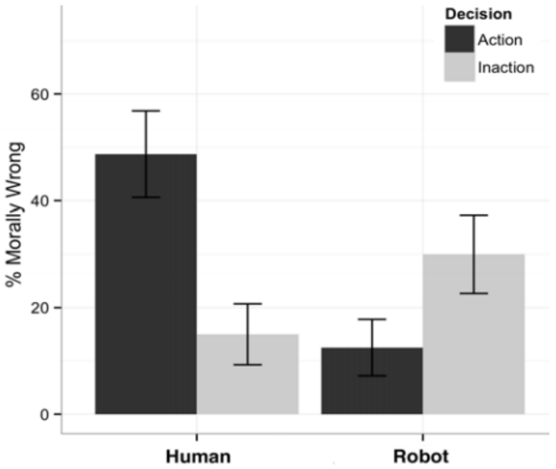

Figure 2. Rates of moral wrongness in Experiment 2 as a function of agent type and the agent's decision

Thus, they considered the robot's intervention as more acceptable than the human's. According to this pattern of results, robots are expected—and possibly obligated—to make utilitarian choices.

3. Methodology

In this study, our aim is to understand how humans react in extreme situations to see if such reactions are applicable to autonomous vehicles. A game, Driver's Dilemma was developed to act as a platform for obtaining data by simulating the Trolley Dilemma. The player will be presented with various scenes. In each scene the player has 3 seconds to decide which direction to steer the car and thus whom to kill. The choices of each player are then studied.

3.1 Trolley Dilemma

The trolley dilemma was first introduced by Philippa Foot [Foot, 1967], and was later built upon by various philosophers by introducing new variants. The trolley problem is intended to test the ethical prowess of people, with its aim being to determine if the person's judgement is utilitarian or not (explained in the next section).

This dilemma is often posed through the following thought experiment. You are the driver of a railway tram, which can only be steered straight or to the left tracks. There are 5 men working, and only 1 man working on the left track. Anyone on the track is bound to be killed and you cannot stop the tram in time. Whom would you kill? Would you take action to sacrifice the one for the many?

Consider another similar scenario. The tram is heading straight towards 5 people and cannot stop. A man is on the foot bridge with another rather large man. The man notices the tram heading towards the people and realises that if he pushes the other large man into the tram's path, the people will be saved and only the heavy man will die. According to utilitarianism, in both cases sacrificing one man's life to save 5 others is equally right.

Foot suggests that we should not think this way. She believed that picking the left track and killing one man is different from pushing a man in front of the track. She believes the latter is morally unjust, insisting that it is one thing to steer towards someone foreseeing that you will kill him, and another to aim at his death as part of your plan.

Our game, Driver's Dilemma, is a simulation of this thought experiment. The player is in a self-driving car which is unable to stop. Throughout the game, the player is presented with multiple scenarios. In each scenario, the car is unable to stop in time and thus the player must decide who should die. The player can choose to crash into a pedestrian who is crossing, a stationary car, a dog, or a group of people, each time different according to the scenario.

We extend the dilemma by adding the option of self-sacrifice for each scenario. In some cases, the player must choose between killing one or a group of people. Other times, the player must simply choose between two different people. The player is encouraged to look at the type of person first. For example, killing a pregnant woman would result in the loss of two lives. Furthermore, in some cases, one option would be to crash into a car. Hitting a car (rather than a pedestrian) will lessen the chance of death of another person, since the other driver may still survive. Lastly, we also included dogs to see if people view dogs as less worthy than human beings.

3.2 Ethical paradigms

Utilitarianism

In this paradigm, actions are seen as morally right if they produce the best possible outcome (i.e. maximize the overall good) [Driver, 2014]. If a person is going to kill a group of people, then it is acceptable to kill the attacker since doing so will save many more lives. This approach is based on impartiality, that is, everyone's happiness is given equal importance.

The means to reach the same consequences do not matter. Thus, using this approach, aiming a gun to his head for a quick and painless death or torturing him till he dies will both be considered acceptable and equally right since the outcome is the same; he dies and the group is saved.

Using this school of thought, utilitarians playing our game should always choose to kill fewer people, regardless of their age or gender. However, when posed with deciding between 2 people, other characteristics must be taken into consideration. For example, killing an old man rather than a kid would be considered morally right using such an approach. This is because the child may bring about more good in his life versus the good the old man can do in his remaining years. We predict that this will be the most popular approach.

Virtue ethics

Virtue ethics emphasizes the role of morality and character in one's choices [Athanassoulis]. If asked to help someone, a virtue ethicist would deem it only moral to give aid. On the other hand, a utilitarian will first assess the situation to determine if helping that person would result in the greater good.

Such ethicists live according to the saying "Do unto others as you would be done by" [Hursthouse and Pettigrove, 2016] and aim to act as a virtuous person would in any situation. They are kind to others not to gain any favours or simply to fulfill some duty, but rather because it is part of their character. Although virtue is derived from natural internal tendencies, it needs to be nurtured to grow. However, virtues and morals are deeply affected by the education received, the moral views of family, friends and society.

When a player is asked to choose between two different types of people, using this approach, their virtues will guide them to make what they consider to be the best decision. For example, for some, killing a dog rather than a human being might be preferred, but such an act is looked down on by others.

Deontology

Deontology is an approach that emphasizes duty and rules [Mastin, 2008]. Rather than focusing on consequences or character, deontologists focus on the morality of the actions themselves. This means that an action is considered right or wrong independently from the situation.

For example, if a doctor was just informed that his family was killed in a car accident by a drunk driver and was asked to perform surgery on the drunk driver, the doctor will be conflicted. On one hand, it is her duty to save the drunk driver, but on the other hand she does not wish to aid the person who killed her family. Deontologists reason that an action is considered moral if it can be made into a maxim, such that it becomes a norm. Thus, they argue that it is the doctor's duty to help the drunk driver. A utilitarian would counter this by saying that by choosing to let the drunk driver die, less people will be killed in the future as he cannot drive under the influence of alcohol again if he does not survive.

People who adopt this approach will choose self-sacrifice since it was their car which failed and thus it is their duty to ensure that no one gets hurt. Even in cases where citizens are not following the law (ex: crossing the road when the lights are red), deontologists will choose to kill themselves as the action of killing someone is always wrong no matter the given circumstance.

3.3 The Social Dilemma

A social dilemma [Tenenberg, 2012] refers to situations where there is a conflict between personal and collective interest. Such a dilemma is applicable to many areas, including self-driving cars. In their paper [Bonnefon et al., 2016], Bonnefon et al concluded that most people would prefer to buy a car that will ensure the safety of its passengers at all costs. However, they would want other people to purchase a utilitarian car even if this means that its passengers might die. Thus, people tend to take the selfish approach for themselves, but do not wish others to do the same.

Due to the social dilemma already being discussed and studied in the paper mentioned above, we did not include such studies in our experiment. The focus of our study is to determine whether the majority deem it acceptable to program autonomous cars using the utilitarian approach.

4. Research Methods

All the data used in this study was obtained from people playing our newly developed game, Driver's Dilemma. The game was put online and shared through social media in order to obtain a sufficient amount of results. The data from the game was sent to a database for storage and evaluation.

4.1 Gameplay

Our game is playable on any computer with an internet connection. The user must navigate to http://gameai3.tk/ and wait for the page to load. Upon loading, the player will be greeted with a menu page and asked to input his/her details (age, nationality and gender).

Basic instructions and background information about the situation of the game are displayed, including the characters used in the game. This is done to ensure that participants will have a good understanding of the rules and choices they will face throughout the game.

After reading the instructions, a test scene is loaded. This scene mimics an actual scene but no data is sent to the database. This allows the player to become familiar with the game before proceeding to the other scenes. There are a total of 10 scenes (excluding the test scene).

The game is set in first person view, displaying the surrounding town from inside a car as though the player is sitting in the driver's seat. The player is warned that the car's brakes are failing. After a few moments, the player will be presented with a scenario in which s/he must decide which direction to steer the car.

These will be scenarios of unavoidable harm, that is, someone will be killed or injured. In each scenario, a person, car or dog are in the direct path of the car. The car does not have enough time to stop safely, and thus the player has to decide whom to crash into. A timer and arrows will be displayed on screen to make the selection. One option will be to crash into a stationary item and kill yourself. If no direction is selected, the car will keep moving straight into the obstacle in front. Upon the completion of these 10 scenes the player may choose to replay the game.

4.2 Design

The main components of this game were obtained from an Asset called Simple Town Lite [Assetstore.unity3d.com A] . This allowed us to easily generate our city. The houses, stores, cars and characters were all imported from this asset. Modifications had to be done to the characters since these were solid game objects and thus were not animated.

We chose to use up-beat music and simple, colourful graphics in order to keep the game light hearted despite the underlying dark situations it portrays. This will help people think more clearly and focus only on the choice ahead rather than the world around them. The graphics were also chosen due to their simple structures and little rendering which is ideal for a web-based game so as to not have lag impeding the user's experience.

Players could exit the game at any time since their choices were sent on completion of each scenario. This ensured that the player did not get bored and select random answers for the scenarios. Thus, the collected results were more reliable and portrayed the true opinions of the participants.

Scenes

1. **Menu scene**: The player is asked to enter their basic details such as gender, age and nationality. This data is then sent to the database and will be linked to all the decisions made by that player. Next some basic instructions are given to the user.

2. **Test scene**: This scene is the first scene which is played. This is used to show the player the type of environment the car will be driving in as well as show them how the scenario and the buttons will look. No data from this scene is sent to the database.

3. The game is composed of 10 different scenes:
   - **Scene 1**: This scene contains a zebra crossing with green traffic lights. Pedestrians cross from both the left and right sides. Law abiding rules come into play since pedestrians are crossing the road while the lights for the vehicles is green.
   - **Scene 2**: This scene is situated at a crossroad. There is a barrier straight ahead, an empty road to the left and a road with a zebra crossing to the right. Pedestrians are walking on a pavement to the left and crossing the zebra crossing on the right. Law abiding rules come into play since the lights are red and pedestrian crossing the road has the right of way.
   - **Scene 3**: Situated right after a corner, the player finds a barrier straight ahead in this scene, while pedestrians are walking on the pavements on both sides of the road.
   - **Scene 4**: This scene is also situated at a crossroad, where the traffic lights facing the player are green. There is a pedestrian walking on the top left pavement and another pedestrian crossing the road without a zebra crossing. Choosing to kill the pedestrian crossing without a zebra crossing would result in killing someone who is not obeying the law.
   - **Scene 5**: After a corner in the road, the player is faced with a barrier straight ahead, while pedestrians are walking on the pavement on both sides of the road.
   - **Scene 6**: In this scene, the player is faced with a zebra crossing with a centre strip in the middle. Pedestrians are crossing the road from both sides. Choosing to go straight ahead will result in the player crashing into the centre strip. Law abiding rules apply since the lights for the player on the left are green. Moving onto the right side of the road would mean coming head on with opposing traffic.
   - **Scene 7**: In this scene, pedestrians are crossing the road without a zebra crossing on both sides, and a lamp post is falling in the middle of the road.
   - **Scene 8**: Here, the pedestrians can be found straight ahead in the middle of the road, as well as crossing from the right side of the road. To the left the player will also find a sign which s/he may choose to crash into.
   - **Scene 9**: In this scene, the player will find themselves faced with a barrier to their left, a broken down car straight ahead with a person inside it, and a pedestrian crossing from the right.
   - **Scene 10**: This scene is similar to Scene 9, with the difference being that the car finds itself in the centre of the scene after it loses control and skids. A pedestrian is found on the right and there's a barrier to the left.

4. **End scene**: Upon completing ten difference scenarios, the final screen is shown. Here, the players are given the option to replay the game or exit.

Key Features in each Scene

1. Outline system: An important visual feature used is the Asset 'Easy Outline' [Assetstore.unity3d.com B]. This feature was used to help distinguish between the objects or characters which the driver can hit. This feature comes with its own script OutlineSystem. It adds a white outline to each game object allowing the user to spot the targets instantly whilst approaching the scenario.
2. Text on characters: In order to identify game objects easily, a label was placed above the targets. A 3D Text Material was used to place these labels above each target, and, with the help of the Text Camera script, the text was set to always face the Main Camera (located within the player’s car).
3. Canvas: In order to display the buttons to the user and guarantee a fixed position on the screen, a Canvas was used. This canvas is made up of three arrows, (left, up, right) which the user can select to steer the car in the direction they would like to go. It also has a timer on top starting at 3 seconds, displayed to the player to show the amount of time left for them to make their decision. The Canvas and all its contents are only visible when the decision process starts.
4. Music: In order to enhance the player’s experience of the game, music was added . [Assetstore.unity3d.com C]. The Music game object is found in the Test Scene since this is the first scene played each time. With the help of the function DontDestroyOnLoad() in the Background Music script, this audio is looped throughout all the scenes without restarting each time.
5. Car: The most important object in the game is the car itself. It was built from scratch and also included a functioning rear view mirror. However, the mirror had to be removed due to its unnecessary heavy workload which slowed the game down when played on a browser. The main camera can be found as a child of the car object, so that it follows it around as it drives. One of the scripts attached to the car is the Crash Script, which is used to trigger the crashing sounds and animation, differentiating between hitting a person and a solid object. The other is the Driving Script which is used to make the car move by following hidden points placed in the scenes. The car also has three different colliders, a box collider used to detect the hitting of an object, a capsule collider used to detect nearby objects, and a sphere collider which is used to activate any movement present in the scene.

Characters

In order to obtain the results for our study, various characters were used in the game. Each character has its own pre-set path. Upon loading the scene, the specific characters placed in the scene are chosen at random with the help of the Scene Controller script. The following list contains all the persons one may come across throughout the game:

- Man
- Woman
- Old Man
- Old Woman
- Boy
- Girl
- Dog
- Pregnant Woman

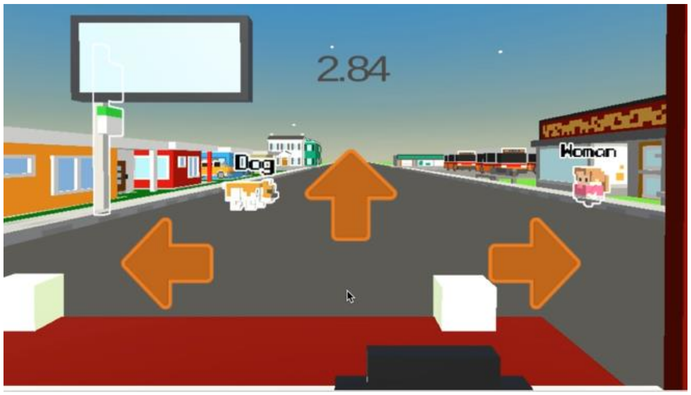

Figure 3: Example of scene with characters outlined and labelled

Furthermore, each person has a small chance of appearing in a group of 5 (all same type of person) instead of alone. Each character is outlined and labelled in the scene on order to ensure that the player is aware of the type of person s/he is about to hit. Thus, the results obtained are more reliable.

Moreover, the scenes of the game are not generated sequentially. These elements of randomness increase the game’s replayability and thus users and more inclined to play more scenes and provide us with more data.

4.3 Collecting Results

Database Implementation

The database was used to store the player’s information as well as their choices as they played the game. The collected data was then evaluated and discussed to produce the statistics on the decisions made, according to the various demographics.

Database Structure

The database contains 3 tables; player_table, choices_table and combinations_table. Their relationships are shown in the diagram below.

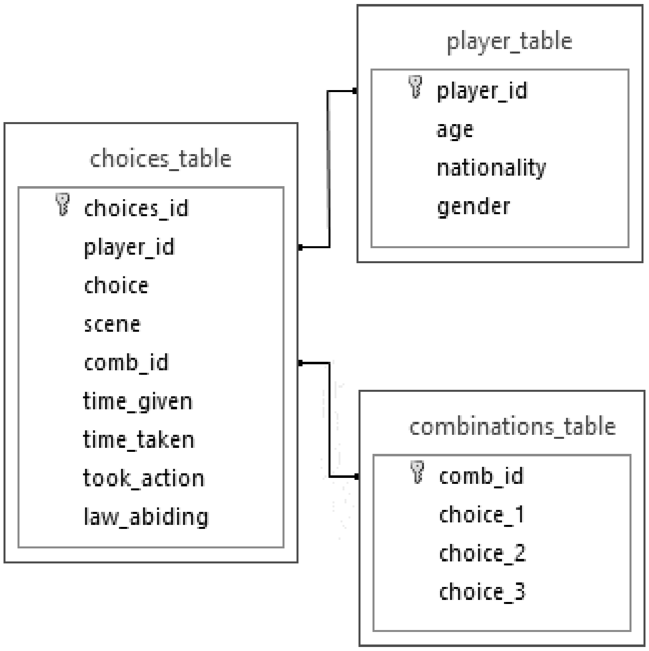

Figure 4: Database Relationships

The player_table contains the information of the player. The player_id is stored in the choices table along with the decision s/he took each time that player finished playing a scene. The choices_table holds information pertaining to the scene the player was faced with, including the choices s/he was given and the decision s/he took. The combinations_table holds all the possible combinations of persons/objects which the player can decide to run over or crash into in each scene.

Inputting into Database

Each new player is asked to fill in their age, gender, and nationality in the menu. Once they press the next button, this data is sent to the player_table in the database.

When a scene loads, the combination of the people/objects in the scene (the choices for who to hit on the left, right and centre) is sent to the database. The combinations_table is checked to see whether that combination already exists, and if not a tuple row is added to the database and the new comb_id (combination ID) is saved to be passed to the choices table once the scene is completed. If the combination already exists in the table, the comb_id of that tuple is saved instead.

Once the player selects an arrow on screen or the timer runs out, the decision that player made is sent to the choices_table. This includes the time given to the player, the time taken to make a decision (will be equal to time given if no action was taken), the player_id, the comb_id and the choice made (ex: Old Man). A boolean variable, called "took_action", is also passed. This will be true if the person made a decision on who to kill and false if no arrow keys or on screen arrows were clicked. This will be used to filter out those people who did not make an active decision.

In scenes where it makes sense, a boolean called "law_abiding" is also passed. This variable will be true when the player hits someone who is disobeying the law. It will be false when the player hits someone who is not doing anything against the rules (ex: person is crossing on a green light).

.Each of these calls are made using SQL to a MySQL database on the server.

Getting Results

The database will be queried once more in order to get subsets from the choices table according to the observation being tested. For example, querying to discover the percentage of people who chose to kill one person instead of five people. Once all results are retrieved, various graphs and other visual aids will be used to display the results.

In total, the game was played by 511 players and a total of 4083 responses were received in just 2 weeks. The discrepancy in the proportionality between the two stems from the fact that not all people played all 10 rounds and some decided to replay the game. 64.19% of participants were males, 33.86% were females and 1.95% identified as other. The mean age was 27 years.

5. Evaluation

In this section, the results obtained from the game and saved in the database are extracted and evaluated. The aim is to determine whether the trolley dilemma would be successful or not in AV cars, while showing the different responses made by participants who choose different directions in reaction to the same traffic accident. From these results a general idea could be drawn based on the majority of the decisions taken by the participants for the same scenario.
The following will be presented:

1. The most killed character/group;
2. The percentage of the participants who took the Utilitarian approach (chose to save the most lives);
3. The percentage of the participants who choose to uphold the law;
4. The percentage of the participants who chose whom to kill according to the age group of the character;
5. The percentage of the participants who chose to commit self sacrifice rather than kill another life
6. Average time taken to decide;

The above results will be divided according to demographic information such as gender, age or nationality as required. The results will determine the preferred approach in such situations and thus, the preferred reasoning autonomous cars should adopt when faced with the Trolley Dilemma.

5.1 Evaluation of Participants

In total, 511 participants played the game, putting in 4083 responses. 64.19% of participants were males (328), putting in 2604 (63.77%) responses, while 33.86% were females (173), putting in 1428 (34.97%) responses. The rest (1.95%, 10 participants), identify themselves with other genders. 19 participants were under 18 years of age, 453 participants fall in the 18-34 age-group, 28 participants were between 35 and 54 years old, and 11 participants were over 55 years of age. 83.17% of participants were Maltese, contributing to 3498 responses, and the rest come from more than 17 other countries around the world, including the United States of America (2.15%), Australia (1.57%), and Canada (1.57%).

5.2 Most Killed/Saved Character

The most killed character was the dog. The dog was presented in 849 scenarios and killed in 53.95% of them. When presented against any other character in the game (Figure 1), including the car passengers, the dog is nearly always chosen to die. The only exception is groups of old women, where, dogs were killed 15.63% of the time and old women were killed in 37.5% of the responses. This clearly portrays the participants' belief that preserving human life is much more important than protecting animals.

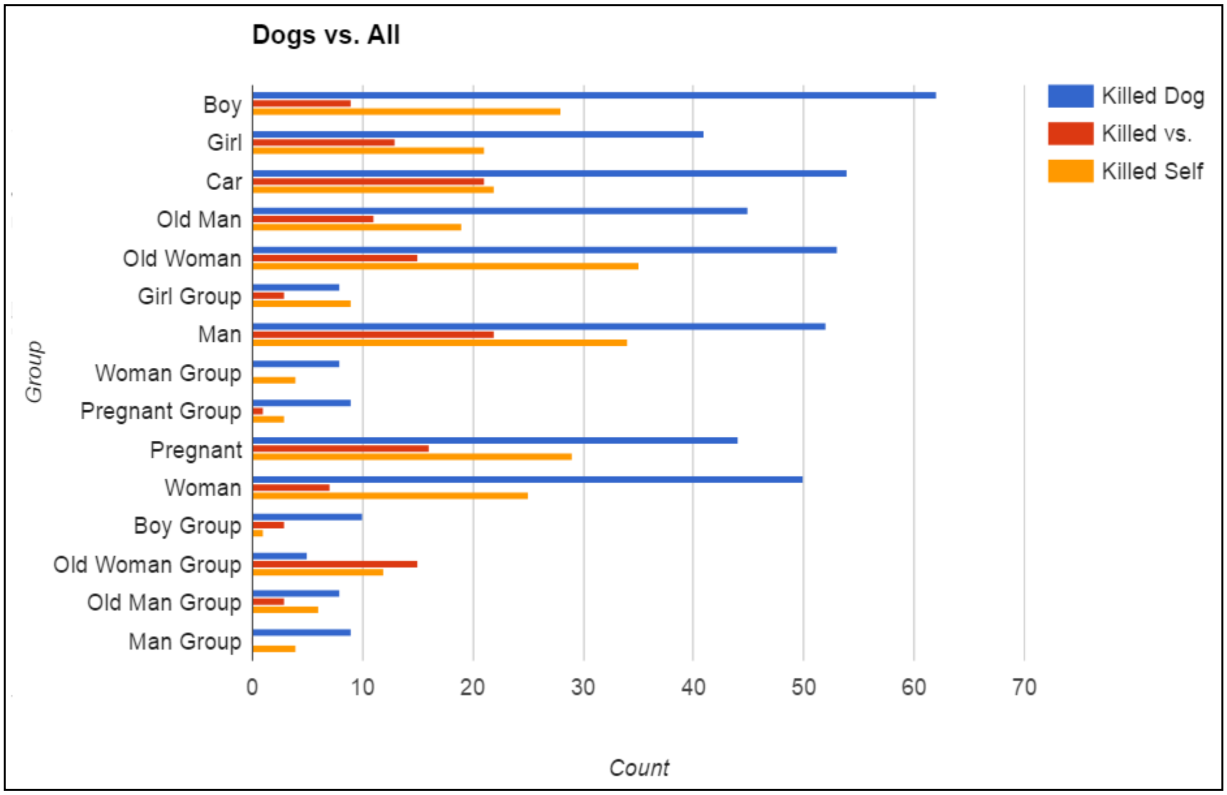

Figure 5: Comparison of the choice to kill the dog/s against each of the other characters

Choice	Count	Kill Count	Percentage
Boy	788	121	15.36%
Boy Group	141	12	8.51%
Car	815	366	44.91%
Dog	849	458	53.95%
Dog Group	106	43	40.57%
Girl	730	120	16.44%
Girl Group	129	16	12.40%
Man	785	194	24.71%
Man Group	136	21	15.44%
Old Man	769	281	36.54%
Old Man Group	143	37	25.87%
Old Woman	825	322	39.03%
Old Woman Group	141	26	18.44%
Pregnant	779	97	12.45%
Pregnant Group	133	12	9.02%
Woman	774	163	21.06%
Woman Group	134	14	10.45%
Self	4083	1779	43.57%

Table 1 – Number of times a character was presented in a scenario and the number of times it was killed

Table 1 shows more valuable information on which characters participants decided to save. Children and adults were killed only 15.04% and 18.28% of the time respectively, whereas old people were killed 35.46% of the time. Gender also plays a small part in the choices made, as males were killed in 24.11% of the scenarios presented, compared to females at 21.12%. It is clearly noticeable that the boy group was given most importance as they were saved 91.49% of the time.

5.3 Discrimination based on number of lives saved (Utilitarian Approach)

From the 933 scenarios in which participants were presented with a group of characters against another single character, only 159 (17.04%) chose to kill groups. The rest either chose to kill themselves as the car passenger (46.84%), or to kill another pedestrian (36.12%) (Figure 6).

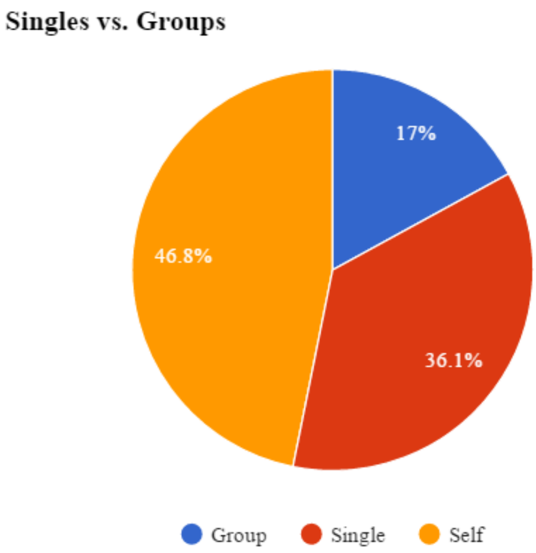

Figure 6: Comparison between the percentage of people who chose to kill a single pedestrian versus themselves versus killing a group

It is clear that killing the least number of lives possible was of utmost importance to the participants, with 82.9% of responses indicating that it would be better if one person is killed (either car passenger or any other pedestrian) rather than a group of people.

This result is comparable to the findings presented by the papers discussed in the scientific review. As seen previously in the study by Navarrete et al [Navarrete et al., 2012], the most preferred action was to maximize the number of lives saved. In their study, the vast majority of participants actively or passively (take no action) choose to sacrifice themselves rather than killing a greater number of pedestrians Similarly, numerous studies mentioned [Bonnefon et al., 2015] [Skulmowski et al., 2012] [Li et al., 2016] highlight the fact that participants choose not to kill groups of people, even at the expense of their own lives

5.4 Upholding the law

In a number of scenarios, participants also had to take into consideration basic road laws and whether characters were breaking or abiding by these laws. Considering only the responses received from these scenarios, which amounted to 1221 (22.9%) responses, not much importance was given to whether the pedestrians were obeying the laws or not. This may be the result of lack of observation from the participants. However, there was still a slight difference, with 5.2% more people choosing to save characters who were abiding the law (waiting on a red light, crossing on a green light, etc.). A similar result can be retrieved when considering solely Maltese participants – in 53.44% of the responses, law abiding characters were saved.

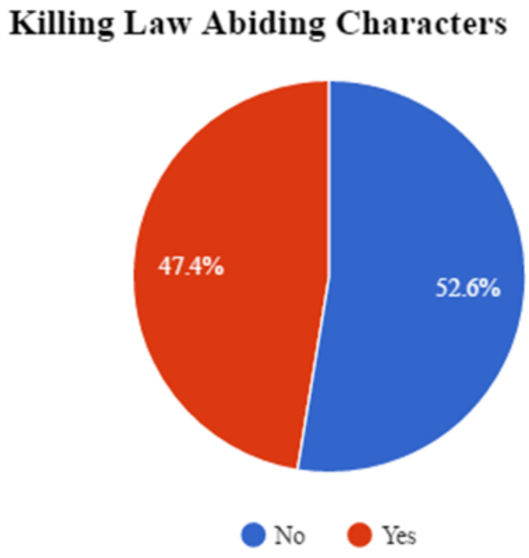

Figure 7: Comparison between number of law abiding and non-lawabiding characters killed (in relevant scenarios only)

Following these results, it is clear that some participants believe that persons who abide the law should be prioritised and protected.

5.5 Discrimination by Age

When comparing the ages of the characters – children, adults, and elderly people – a lot of information can be retrieved. Participants tended to prefer saving young children (Figures 8 and 9), where 17.6% were killed when presented against adults, and only 9.8% were killed against elderly people. Additionally, the trend of saving the young continues at higher ages – only 15. 8% of participants chose to kill an adult when presented against elderly people.

It is clear that participants prefer to kill older people, while prioritizing the protection of young children. Presumably, these participants would want autonomous vehicles to somehow determine the age of the people in a scene and save the youngest people when such a scenario occurs.

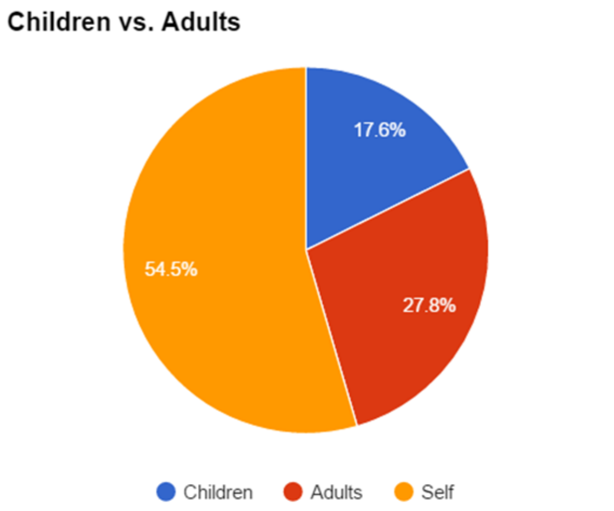

Figure 8: Comparing percentage of children versus adults hit

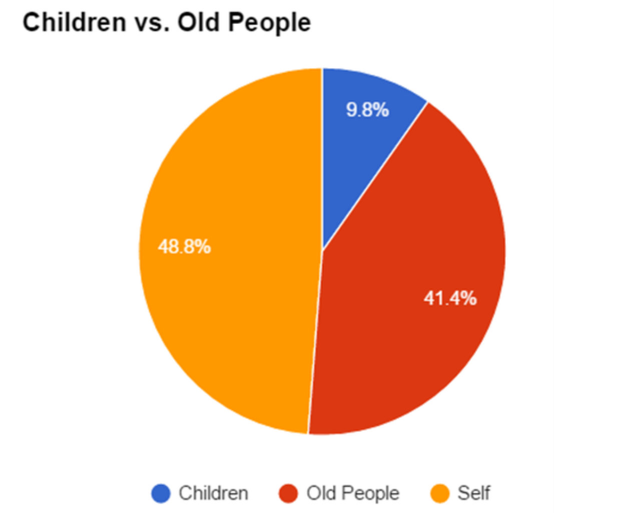

Figure 9: Comparing percentage of children versus elderly hit

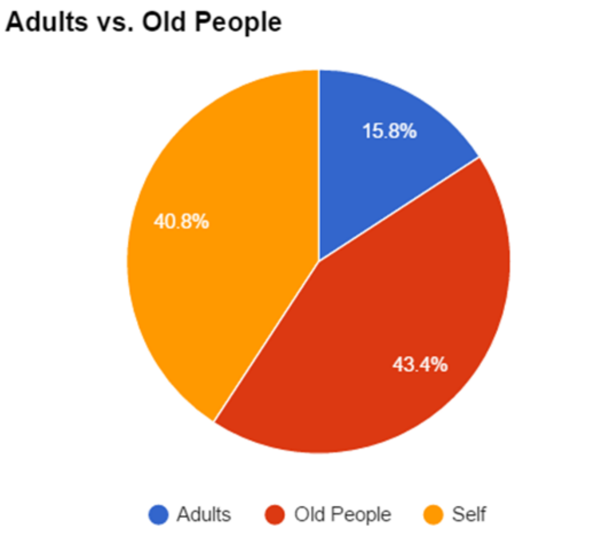

Figure 10: Comparing percentage of adults versus elderly hit

5.6 Time taken to make Decision

The average time taken to make a decision was approximately 1.36s, where participants were given 3s. This figure also includes the 401 responses in which no action was taken (which amount to 9.8% of the responses), for which time taken was recorded to be 3s. Considering only responses in which action was taken, the average time taken drops to approximately 1.18s. 15.7% of the responses arrived in the last available second (Figure 11). An overwhelming majority (74.5%) of participants assessed the scenario presented and took action in less than 2 seconds.

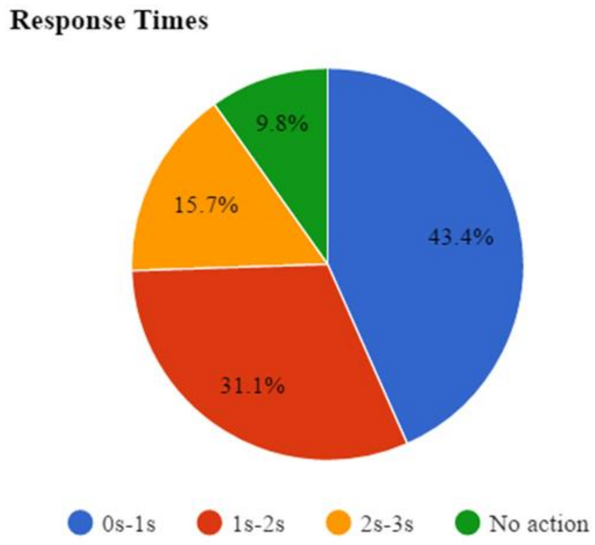

Figure 11: Response Times

The results show that participants were given enough time to make their decision; with the majority taking less than one second to make their choice. However, this may also result in some inaccurate results due to a lack of observation.

Interestingly, the average time take for a decision increases slightly becoming 1.41s when a group of people is presented in the scenario. This implies that the addition of a group of people did not have a noticeable impact on the amount of time taken to make a decision. This contrasts with the study carried out by Skulmowski et al. [Skulmowski et al., 2012] which concluded that when presented with a group of characters, response was faster suggesting that Utilitarian decisions can be made in a shorter amount of time.

These results help in determining whether autonomous vehicles should ask the driver to take a decision when the scenarios presented by simulation occur in real-life.

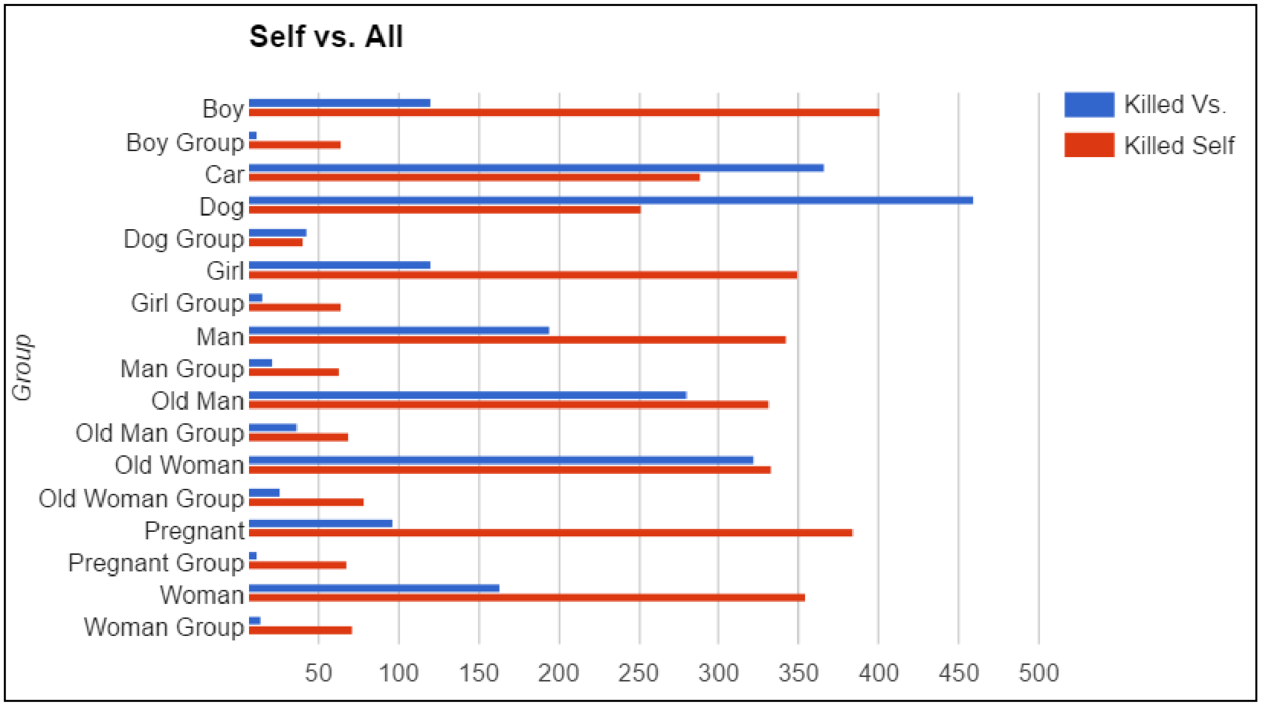

Figure 12: Comparison of the choice to kill oneself against each of the other characters

5.7 Killing Self vs. Killing Others

From 4083 responses, 43.57% of participants decided to kill the driver in the self-driving car, which represents the participants themselves.

One of the factors that could have played into this decision was the fact that the accidents were partly the driver's fault, since their own car's brakes were failing in these scenarios. Maltese participants opted to kill themselves in 42.31% of the case. Noteworthy is the comparison between genders – the majority of male participants, with 52.37%, chose to kill themselves, against the 47.34% of female participants and the 43.14% of participants who identify themselves as other genders.(Figure 12)

When evaluating which characters participants decided to sacrifice themselves for (Figure 7), it emerges that participants would save all human characters rather than save themselves. However, when presented with dogs, only 36.9% of participants decided to kill themselves. Furthermore, 55.9% of participants decided that it is better to hit a car (killing its passenger), rather than kill themselves. When presented with a group of characters, an overwhelming majority of 74.22% decided to save the group.

5.8 Evaluation Summary

Through the results presented, the most important decisions according to the participants can be listed as follows (highest priority at top):

1. If animals are present, kill them – participants prefer to kill a dog rather than any other character in the simulation
2. Hit another car if possible (possibly killing the car's passenger) – nearly 45% of participants decided to hit a car rather than killing themselves or other pedestrians
3. Save groups of people – 82.9% of responses prefer killing a single person (car passenger or pedestrian) rather than groups of people
4. Sacrifice oneself rather than killing other people – participants prefer to kill themselves rather than killing any other character (except dogs and other cars)
5. Save younger people – participants prefer to kill elderly people rather than adults or children
6. Save law abiding people – around 53% of participants saved law-abiding characters

6. Conclusion

With the use of a web-based simulation, a study was conducted on people's thoughts on car accidents where lives are at stake. This was done in order to conclude how autonomous vehicles should react in such situations.

From the results, it was clear that saving human lives is of utmost importance to people. The life of a dog was seen as less valuable in most scenarios. Whenever possible a crash with another car, where the other driver has a chance of surviving, was more favourable rather than running over a pedestrian. Although this was quite expected, it is still an important to note that autonomous vehicles should aim to prioritize human lives over anything else in cases of unavoidable injury. Our study could be built upon by providing the self-driving vehicle with a mechanism to calculate the probability that the other car driver or pedestrian survives the crash. This would ensure that it is truly the safety of the people involved which is being taken into account. It is also expected if other animals are used in the scenarios, they will be killed more often since dogs are seen as pets and are friendly to humans.

Continuing on the point of saving human lives, the utilitarian approach was the most common methodology used among participants. The number of lives which were saved had a significant effect on the choice of the participants, with many opting to kill the few instead. It is valuable to note that there were a number of different variables which were analysed, and saving the highest number of lives was one the most influential factor out of all of them, with only 17% choosing the kill the many. It should be noted that the groups in this simulation were represented by five people at once, and that the noted results may vary depending on the number of people within these groups.

The age of the lives being saved was also a notable factor. The younger the person, the more likely that they were to be saved. This choice usually follows the reasoning that the young still have many years left and thus a lot to experience in comparison to the older counterparts. Interestingly, pregnant women were the most saved character when groups weren't involved. This further confirms the previous points, since saving a pregnant woman can be viewed as saving two lives at once, with one of them being of a very young age.

Whether the other people in the scenarios were following the law or not was also taken into account, albeit briefly. However, the difference was slight and thus it might stem from the fact that some participants did not observe the scene well and thus did not realise such details. A further study could make the laws which people are following or disobeying more obvious to the participants, to get a better indication of how much it changes their perception.

An option which was explored in this study is the ability to sacrifice your own life in order to save others. Quite a large percentage of people chose that option when they were faced with the alternative of killing another person. This shows that not all people would want to ensure the safety of the driver at all costs, and might want autonomous cars to also give priority to other lives as well. However, it must be noted that in these scenarios it was the driver's own car which was at fault, and therefore further studies could highlight what's people's choices would be depending on who caused the accident.

These results shed a great light on how people think that autonomous cars should handle such scenarios. Further research could determine whether participants would want to buy or use autonomous vehicles, and the extent of autonomy that they would want such cars to have. Taking into consideration the presented scenarios, would self-driving car users want their car to take the final decision, or would they prefer if the car waits for the user's decision? What happens if the user does not make a decision in time?

Another possibility which can be explored in the future in extension to the last point, is the fact that there can also be other passengers in the car with the driver. Depending on whether a family member or a stranger is involved, the rates of sacrificing themselves or the lives of people outside the car are also expected to change. Also, the amount of immersion of the simulation can be improved with the use of Virtual Reality instead of on a screen, in order to make the participants experience the emotions typically evoked by such situations.

7. References

[Bonnefon et al., 2016] Jean-François Bonnefon, Shariff Azim, and Rahwan Iyad. *The social dilemma of autonomous vehicles*. Science 352.6293: 1573-1576, 2016.

[Bonnefon et al., 2015] Jean-François Bonnefon, Shariff Azim, and Rahwan Iyad., *Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?*, 2015.

[Foot, 1967] Philippa Foot. *The problem of abortion and the doctrine of double effect.*, 1967.

[Driver, 2014] Julia Driver. *The History of Utilitarianism,* The Stanford Encyclopedia of Philosophy, 2014.

[Hursthouse and Pettigrove, 2016] Rosalind Hursthouse and Glen Pettigrove. *Virtue Ethics,* The Stanford Encyclopedia of Philosophy, 2016.

[Athanassoulis] Nafsika Athanassoulis. *Virtue Ethics*, Keele University

[Mastin, 2008] Luke Mastin, *Deontology*, The Basics of Philosophy, 2008.

[Tenenberg, 2012] Josh Tenenberg, University of Washington,Social Dilemma and related Definitions, 2012.

[Navarrete et al., 2012] C. David Navarrete, Melissa M. McDonald, Michael L. Mott, and Benjamin Asher. Virtual Morality: Emotion and Action in a Simulated Three-Dimensional "Trolley Problem". 2012

[Skulmowski et al., 2012] Alexander Skulmowski, Andreas Bunge, Kai Kaspar and Gordon Pipa, Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study, Frontiers in Behavioral Neuroscience, 2014.

[Malle et al, 2015] B.F. Malle, M. Scheutz, T. Arnold, J. Voiklis and C.Cusimano, *Sacrifice One For the Good of Many?: People Apply Different Moral Norms to Human and Robot Agents*. In Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction,2015

[Li et al, 2016] Li, J., Zhao, X., Cho, M. J., Ju, W., and Malle, B. F. *From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars*. SAE Technical Paper, 2016

[Waymo] *On the Road – Waymo*. *Waymo*. Retrieved from https://waymo.com/ontheroad/

[Assetstore.unity3d.com A] *A: Simple Town Lite – Cartoon Assets Asset Store*. *Assetstore.unity3d.com*. Retrieved from https://www.assetstore.unity3d.com/en/#!/content/43480

[Assetstore.unity3d.com B] *B: Easy Outltine Asset Store*. *Assetstore.unity3d.com*. Retrieved from https://www.assetstore.unity3d.com/en/#!/content/62384

[Assetstore.unity3d.com C] *C: Cartoony Background Music Asset Store*. *Assetstore.unity3d.com*. Retrieved from https://www.assetstore.unity3d.com/en/#!/content/16756