Ethics of Autonomous Vehicles: Australians' Expectations and Moral Preferences

Amir Rafiee[a]
School of Information and Communication Technology, Griffith University, Brisbane, Australia
fi *fi*

Keywords: Autonomous Vehicle, AI Policy, AI Ethics, Automated Driving System, AI Regulation, Self-Driving Cars, Ethical Dilemmas.

Abstract: Autonomous Vehicles (AVs) can handle most driving scenarios, but ensuring safety in every situation remains a challenge. Factors such as technology failures, faulty sensors, and adverse weather introduce complex ethical dilemmas that AVs must navigate. Considering the societal benefits of AVs, it is crucial to address both technical challenges and ethical expectations. This paper evaluates Australians' perceptions and expectations regarding the ethical programming of personal AVs in six dilemma scenarios using a structured questionnaire. The participants selected the most acceptable outcome in each scenario, informed by ethical and legal considerations. The survey offers a framework for understanding public moral preferences by excluding discriminatory factors and considering legal contexts. The findings prioritise Australians' preferences for ethical AV behaviour, focusing on Injury Over Sacrifice (IOS), Harm Confinement and Lawfulness (HCL), and Harm Prevention and Prioritisation (HPP). These insights can guide policymakers and manufacturers in aligning AV programming with societal values. The study also highlights how ethical models like the Objective Decision System (ODS), which selects outcomes randomly when no clear moral preference emerges, can balance public trust and responsibility in AVs.

1 INTRODUCTION

Autonomous Vehicles (AVs) can manage most driving scenarios with relative ease; however, designing a system that ensures safety in every situation remains highly challenging (Campbell et al., 2010). For example, recognising humans and other objects on the road is both critical and more difficult for AVs than for human drivers (Farhadi et al., 2009). Consequently, in future AVs, crash avoidance features alone will not suffice. It may not always be possible to avoid an accident, especially as AVs navigate city streets and avenues, which are more dynamic than highways. Factors such as technology errors, faulty sensors, malicious actors, and bad weather can contribute to inevitable collisions (Gomez et al., 2014). Given the risks associated with fully autonomous vehicles, it is clear that they require a set of principles to govern their utility. Moreover, for individuals to trust AVs, their design must align with ethical and inclusive values. This has led to a global response, with nations addressing the emerging ethical issues surrounding AI-enabled technologies.

The major challenge, however, extends beyond ethical dilemmas. If we assume that the widespread adoption of AVs is necessary for the societal benefits these vehicles are expected to provide (Bonnefon et al., 2016), it is evident that any reasonable ethics policy should also consider the various expectations of users and the larger society in which they are implemented. This necessitates ensuring user satisfaction and safety, as well as other essential design values such as trust, accountability, and transparency (IEEE, 2016). Furthermore, ethical theories and people's expectations often overlap, even if they will always be somewhat in tension. Therefore, it appears that an ideal ethics policy must resolve this tension by balancing public acceptability and moral requirements. It must be acceptable enough to gain trust and adoption from users while remaining moral enough to avoid reflecting negative human tendencies (Shahriari and Shahriari, 2017).

Some scholars have used polling and surveys to understand public opinions on these moral issues (Awad et al., 2018; Bonnefon et al., 2016). Although these polls can assist in developing moral algorithms, they cannot determine their implementation. There-

[a] https://orcid.org/0000-0002-8149-4780

Rafiee, A.
Ethics of Autonomous Vehicles: Australians' Expectations and Moral Preferences.
DOI: 10.5220/0013123900003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 1*, pages 287-297
ISBN: 978-989-758-737-5; ISSN: 2184-433X

fore, while public polling on how AVs should handle dilemma scenarios can inform regulation, there is a compelling argument that the final decision should be left to experts (Bonnefon et al., 2015). Nonetheless, a consensus is necessary, and important ethical decisions should not be left solely to engineers or ethicists. Algorithms that do not align with societal moral expectations or consumer preferences are likely to hinder the smooth adoption of AVs. Consequently, any solution for programming moral dilemmas in AVs should not discourage potential buyers or the public at large.

In this paper, I aimed to evaluate Australians' perception and expectation regarding personal AVs relating to various ethical settings. Section 2 contains the definitions and considerations which will be used throughout the paper. In Section 3, I critically assessed the Moral Machine (MM) experiment, which aimed to establish a global representation of people's moral preferences in various AV dilemma scenarios. I evaluated this study and explained why certain dilemma scenarios and participant options should be excluded from future research. In Section 4, I argued for the inclusion of random selection as an option in the survey questionnaire. I discussed the data collection method in Section 5, followed by an analysis of the data and findings in Section 6. My proposed ethical policy and AV ethics framework were presented in Section 7. In conclusion, I revisited the primary research objective and summarised the findings, aiming to propose a publicly acceptable and ethically sound framework for programming AVs.

2 DEFINITIONS AND CONSIDERATIONS

In this section, I provide the necessary definitions and considerations that form the foundation for the remainder of the paper. The aim was to clarify the key concepts and frameworks discussed throughout, ensuring a common understanding of the topic at hand.

2.1 Autonomous Vehicles

There are a wide range of AVs, which supplement or replace human drivers with AI. Meaning that a human driver either has limited responsibility or does not need to be present at all. The Society of Automotive Engineers (SAE) defines 6 levels of automation that apply to automotive vehicles. Levels 0–3 require that human drivers take control of various vehicle operations under certain conditions, and levels 4 and 5 have no such requirements (Shadrin and Ivanova, 2019). In this paper, AVs refers to autonomous vehicles at level 4 or 5 of driving automation, as defined by the SAE. These are vehicles where no driver is ever needed, or there might be an option for human override but not a requirement.

3 RESEARCH QUESTION DESIGN

There were six dilemma situations shown in the survey questionnaire involving an AV, and the participants were asked to decide which outcome was the most acceptable to them. The purpose of the survey questionnaire was to evaluate respondent's perception and expectation on personal AVs relating to various ethical settings.

I also proposed and included a theoretical solution called an objective decision system (ODS). If selected by a participant, the AV would consider all feasible and justified outcomes for a given crash scenario, based on a philosophical or moral framework, and choose one at random. Numerous alternatives were excluded because they were either unjustifiable or discriminatory. In the questionnaire, I asked participants five demographic questions, including gender, age, education level, their perceived consumer adopter category, and whether they currently own a Level 2 AV. Adopter categories divides consumers based on their willingness to adopt new innovations. Understanding these categories helps in analysing preferences toward AV crash scenarios. Using the survey responses and the demographic variables, I then compared people's preferences towards a pre-determined outcome based on the different crash scenarios.

3.1 Exclusion Criteria

The MM Experiment, surveyed people across hundreds of countries to gauge moral preferences in AVs and what priorities they should have in the event of an unavoidable accident. The researchers used an online survey to collect over 39 million responses to hypothetical ethical dilemmas for AVs (Awad et al., 2018). The strongest preferences were for sparing human lives over animal lives, sparing more lives, and sparing young lives. There was a general indication by the results, that there was more preference for sparing children's lives, rather than those of adults. Notably people from different parts of the world had dissimilar beliefs on how AVs should make such life and death decisions.

3.1.1 Human Life vs Animal Life Decisions

One of the many crash scenarios in the MM experiment involving an AV, presented participants with a choice between saving a human life or an animal. Generally, when an animal appears directly in front of a vehicle and there is not enough time to brake or swerve (with road conditions being a factor), the safest option is to continue ahead and strike the animal (Curtis and Hedlund, 2005). Currently, drivers are not legally liable if they hit a wild animal, such as deer, elk, and other smaller animals. It can be assumed that this would also apply to AVs. Additionally, German Ethical Rule 7 clearly states that in dilemma situations, the protection of human life should have top priority over the protection of animal life. This rule aligns with social expectations assessed through the MM experiment (Luetge, 2017). Therefore, I excluded similar scenarios from the study, as it is apparent that in an imminent crash where an AV must choose between saving an animal life or a human life, the human life should take precedence.

3.1.2 Discriminatory and Immoral Decisions

Another crash scenario in the MM experiment involved an AV faced with an imminent collision, requiring participants to choose between saving an elderly or a young person. The justification of discrimination is crucial when analysing trolley-problem-like scenarios. Philosophers distinguish between non-arbitrary discrimination, which is morally justified, and arbitrary discrimination, which is not (Rachels and Rachels, 2012). For example, excluding blind applicants from air traffic control jobs is non-arbitrary because visual acuity is essential for safety. In contrast, excluding individuals based on age, race, or gender without relevant justification is arbitrary and unjust. In crash scenarios, age-based discrimination is morally equivalent to other forms of arbitrary discrimination, such as those based on race or gender (Lin, 2013). Therefore, discriminatory scenarios have been excluded from this survey research.

3.2 Inclusion Criteria

3.2.1 Legal Implications

The MM experiment, presented as an applied trolley problem based on Thompson's case, focused solely on moral responsibility while neglecting the legal issues that could significantly impact participants' decisions and constraints on their rights to action (Etienne, 2020). As observed in other experiments, real conditions can influence participants' decisions (Francis et al., 2017). People's moral choices often fluctuate and are heavily affected by the amount of information available and the degree of deliberation involved. Many participants might contend that, with more information, their decisions could differ (Noothigattu et al., 2018). Therefore, in the survey research, I presented the legal implications of each of the six dilemma scenarios for participants to consider, ensuring their decisions were informed by both ethical and legal perspectives.

4 THE CASE FOR RANDOM SELECTION

4.1 Introduction

In this section, I draw upon existing literature to present arguments supporting the use of random selection in future AVs, particularly in dilemma situations where, from a legal perspective, all available outcomes are justifiable on philosophical or moral grounds. I contend that lotteries can eliminate bias, maintain ethical standards by excluding immoral or illegal options, and ensure transparent, tamper-proof outcomes. Decision-making by lot is likely to be simple, objective, and cost-effective, making it a serious consideration for researchers, AV manufacturers, and policymakers. This approach ensures that AVs navigate ethical challenges justly and equitably, promoting trust and acceptance of AV technology.

4.2 Random Selection Applied to AVs

Imagine an AV navigating a busy intersection when a pedestrian suddenly steps into the road. The AV faces two immediate options, brake to avoid hitting the pedestrian, potentially causing a collision with a closely following motorcyclist; or continue straight, hitting the pedestrian but avoiding harm to the motorcyclist (Coca-Vila, 2018). In this scenario, both options have valid moral justifications: prioritising pedestrian safety versus preventing harm to the motorcyclist. Rational analysis based on safety metrics and traffic conditions might fail to determine a clear superior option. Additionally, from a normative viewpoint, they constitute a symmetrical community of danger in which saving the life of one is only possible at the cost of the life of the other. According to criminal law, when two duties to act are in conflict, the self-driving car has to decide which interest to protect and, as such, it will always be acting according to the law (Zimmermann, 2014). This means that, in

the same way that a father is permitted to breach the obligation to save one child from drowning by saving the other, the self-driving car in the example also has to decide who will suffer the harm awaiting both victims, without causing a transfer of harm to an innocent third party (Coca-Vila, 2018). In short, braking the car to prevent running over the pedestrian is not, therefore, unlawful homicide, and neither would be not stopping and hitting the pedestrian crossing the road in front of the car. If both possible outcomes are acceptable, then the car can be programmed to decide through a random decision system (RDS) (Zimmermann, 2014). Consequently, the AV ethics setting activates its RDS to decide whether to brake or continue straight.

One of the primary benefits of using a randomised decision in AVs is its ability to neutralise biases. A random decision system eliminates these biases by making the choice impartially, solely based on chance rather than subjective factors. Critics might argue that using a random decision system bypasses rational thought and deliberation, potentially leading to suboptimal outcomes; however, where no decision-making parameter exists and the alternative of not making a decision has been rejected, randomness is seen as the best solution to solve undecidable conflicts (Rescher, 1960).

First, employing a lottery-based decision-making process ensures that the outcome is not influenced by arbitrary or subjective factors. This approach upholds fairness and prevents the vehicle from making biased decisions based on irrelevant considerations (Dworkin, 2011). Secondly, random decision systems in AVs plays a critical role in preserving ethical standards. By excluding immoral or illegal options from consideration and focusing solely on morally acceptable choices, AVs uphold ethical norms and legal regulations. This ensures that decisions align with societal values and legal frameworks, promoting trust and acceptance of AV technology among stakeholders and the general public.

Thirdly, the fact that the car makes a random decision and automatically acts on it removes any risk of manipulation, both in the choice and implementation of the solution. Therefore, the victim will know that all outcomes are considered equally, and his faith is resting in the hand of an objective force (Coca-Vila, 2018). Duxbury (1999) also argues that a decision made by lot, offers a fair way of dealing with many uncomfortable, or even inherently unfair, dilemmas. A non-biased lottery would remain impartial and thus less susceptible to corruption (Duxbury, 1999). Finally, decisions determined through random selection lack a human element, suggesting that accountability for challenging choices can be alleviated from specific individuals or groups (Duxbury, 1999). Moreover, a non-random system may be more susceptible to manipulation. For instance, if pedestrians believe the system always favours them, they might become careless. Even more concerning, malicious actors could exploit this predictability to stop vehicles for their own purposes. By eliminating complete predictability, a random system discourages both reckless behaviour and exploitation.

5 DATA COLLECTION METHOD

This research employed a descriptive survey design to effectively gather original data from a population too large to observe directly (Botes, 1996). The descriptive survey design aligns with the study's objective, which aimed to assess respondents' perceptions and expectations regarding personal AVs in various ethical settings. Participants were drawn from the general Australian population aged 18 and above, with Australia's diverse demographics, including its status as the country with the highest immigrant proportion among nations with populations over 10 million, ensuring a representative sample. The Australian population was 25 million as of December 31, 2020 (Australian Bureau of Statistics, 2021). Data collection was conducted via SurveyMonkey, targeting eligible respondents across various age categories. The platform stored responses securely, and the data was subsequently analysed using SPSS.

The chi-square analysis of variance was chosen as the statistical method to evaluate the survey questionnaire results due to its suitability for examining relationships between categorical variables, which aligned with the research questions posed. The confidence level was set at ninety-five percent to ensure robustness in the conclusions drawn from the data. SPSS software was used, specifically employing the cross tabulation feature under Descriptive Statistics to generate a contingency table displaying frequencies and column percentages. Detailed statistical analyses and accompanying charts are available upon request; however, due to word limit constraints, they have been excluded from this submission.

6 DATA ANALYSIS AND FINDINGS

6.1 Overview

To sufficiently understand the results, readers are advised to familiarise themselves with the scenarios outlined in the survey questionnaire. The following provides a summary of the 6 AV crash scenarios, the associated moral dilemmas, and the legal implications, offering the necessary context to understand the results discussed in this section. As noted in Section 3, a theoretical solution called the ODS was proposed and included. If selected by a participant, the AV would consider all feasible and justified outcomes for a given crash scenario based on a philosophical or moral framework and choose one at random.

Scenario 1 - Brakes Failure on Mountain Road
The AV faces a dilemma where its brakes have failed, and it must decide between swerving right to hit a bystander on the mountain or continuing straight off a cliff, resulting in the death of the passenger. Legally, the bystander is not responsible for the emergency situation or the AV's mechanical failure, which means the AV cannot legally prioritise saving its passenger by causing harm to the bystander.

Figure 1: The ratio between AV passengers and bystanders was equal, with one passenger for every bystander.

Scenario 2 - Intersection with Illegally Crossing Pedestrians
An AV with one passenger encounters a dilemma at an intersection where five pedestrians are illegally crossing the road. The AV has two options: it can either swerve left to avoid the pedestrians, potentially killing a cyclist, or continue straight, resulting in the likely injury or death of one or more of the pedestrians. The AV's passenger will remain unharmed in either case. Legally, the pedestrians are at fault due to their illegal crossing, and the AV could justify its decision to continue straight as a defensive emergency, meaning it would not be breaking the law by prioritising the cyclist's safety over that of the pedestrians who caused the dangerous situation.

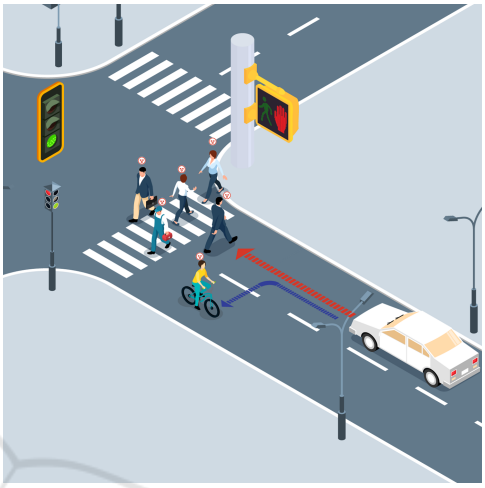

Figure 2: Five times as many people illegally crossed the road compared to the innocent cyclist. The AV passenger remains unharmed regardless of the outcome.

Scenario 3 - Highway Dilemma with Fallen Pallet
The AV with a single passenger encounters a highway dilemma when a pallet falls from the truck ahead. The AV has three options: continue straight, resulting in the passenger's death from colliding with the pallet; swerve right, leading to the death of a helmetless motorcyclist; or swerve left, causing severe injury but survival of a fully geared motorcyclist, with no harm to the passenger in these latter scenarios. The helmetless motorcyclist's illegal activity doesn't prevent his death if hit, while the fully geared motorcyclist may suffer bone fractures but can seek compensation. This scenario poses a moral dilemma between prioritising preserving life versus avoiding physical harm.

Scenario 4 - Pedestrian vs. Motorcyclist
The AV faces a dilemma where it must decide between two potentially fatal outcomes involving a pedestrian and a motorcyclist. If the AV brakes to save the pedestrian who has suddenly appeared in its path, the motorcyclist following closely behind may collide with the AV and be killed. Alternatively, the AV could continue on its course, striking the pedestrian and saving the motorcyclist. Importantly, the passenger in the AV remains unharmed in either scenario. From a legal perspective, both the pedestrian and the motorcyclist could be considered negligent

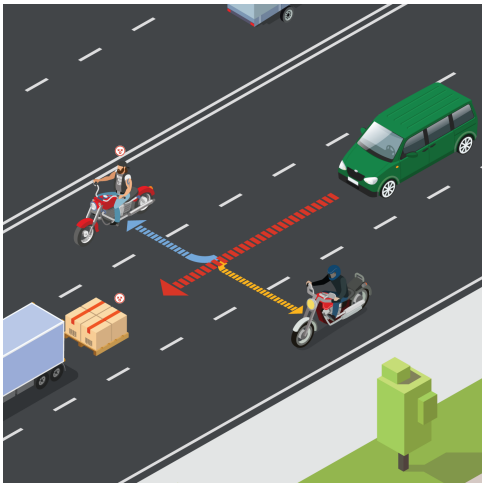

Figure 3: The distribution between AV passenger and motorcyclists was even, indicating that in all three possible outcomes, either one passenger or one of the motorcyclists would be harmed.

for their actions, which contributed to the situation. However, the AV and its passenger would not bear legal responsibility for the outcomes because the AV took reasonable measures to avoid foreseeable harm, thus absolving them of liability under the current legal frameworks.

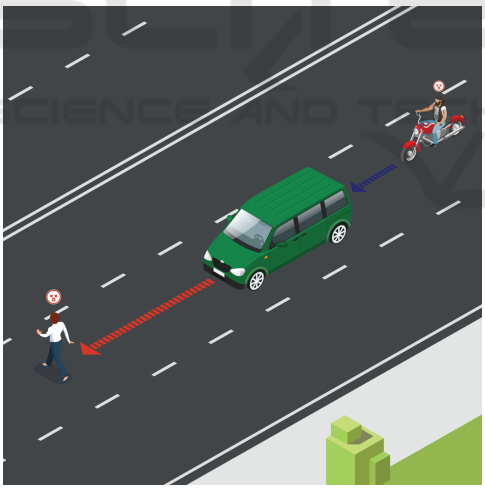

Figure 4: The ratio between the negligent pedestrian and motorcyclist was equal, at one to one. The AV passenger remains unharmed regardless of the outcome.

Scenario 5 - Intersection with Legally Crossing Pedestrians

The AV encounters an ethical dilemma at an intersection where five pedestrians are crossing legally with a green light. The malfunctioning AV must decide between swerving into a wall, risking the lives of its five passengers to save the pedestrians, or proceeding straight, potentially endangering one or more pedestrians to protect its passengers. Legally, the pedestrians are in the right as they cross with the green light, implying no fault on their part. This legal framework highlights the ethical challenge posed by AV technology, where programming decisions must conform to legal norms while addressing intricate moral dilemmas in unpredictable situations.

Figure 5: The number of pedestrians legally crossing the road equalled the number of AV passengers illegally crossing the intersection, with both totalling five instances each.

Scenario 6 - Intersection with Legal Pedestrian Crossing

The AV faces a critical decision during loss of control at an intersection. It can either swerve to the right, colliding with a tree and resulting in the deaths of three passengers, or continue straight through the intersection, striking a pedestrian legally crossing on the zebra lines, thereby saving the passengers. The legal implication centres on the pedestrian's right of way, as they are crossing legally at the zebra lines, prioritising pedestrian safety over vehicular movement. This legal context adds complexity to the ethical dilemma faced by the AV, highlighting the potential clash between moral imperatives and legal responsibilities in autonomous driving scenarios.

6.2 Discussion

In this section I summarised the participants' moral preferences, which could be used to develop an ethical policy for AV decision-making. I then compared these preferences to the three main philosophical theories which have been extensively discussed in the context of AVs, specifically utilitarianism, deontological ethics and the Rawlsian approach, and evaluated which framework best aligns with the proposed eth-

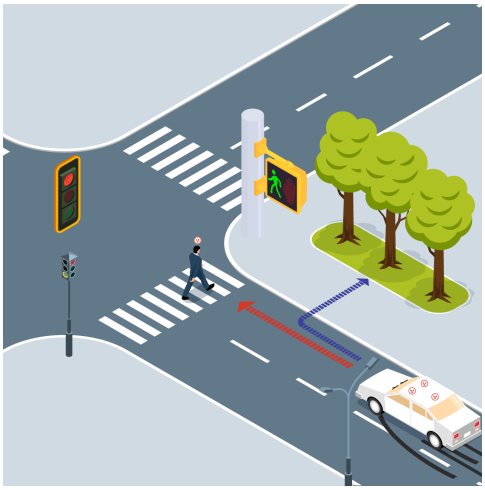

Figure 6: The ratio of AV passengers illegally crossing the intersection compared to pedestrians legally crossing the road was 1 to 3.

ical principles for AV programming (Rafiee et al., 2023).

6.2.1 Australians Moral Preferences

Objective Decision System (ODS). Overall, the ODS was the most preferred outcome among all participants. The use of the ODS could be justified as it provides a straightforward and comprehensible solution that appeals to a broader audience. It offers a common-sense approach that resonates with a larger population. This approach could be particularly useful in the early development stages of AVs at levels 4 or 5 of driving automation, as discussed in Section 2, where no driver is needed, or there might be an option for human override but not a requirement. These AVs may still lack the technological capabilities to quickly assess all viable outcomes and make a split-second decision in a moral dilemma involving potential loss of life. Such technological limitations could delay their deployment and adoption.

Injury Over Sacrifice (IOS). The only crash scenario which did not receive the highest vote for the ODS, was question 3 where the majority of the participants opted for the AV to swerve to the left, hitting and severely injuring the motorcyclist with a helmet. This choice reflects a prioritisation of minimising fatalities, with severe injury seen as a more acceptable outcome than death. The IOS rule can be programmed into an AV from the perspectives of utilitarianism, deontology, and Rawlsian ethics. Utilitarianism supports IOS as it minimises overall harm by opting for injury over the loss of life. Deontological ethics aligns with IOS by prioritising the duty not to kill, thus making the rule morally acceptable. Rawlsian ethics, focusing on fairness and justice, also supports IOS as it respects the fundamental right to life and aims to minimise the worst possible outcomes. Therefore, IOS can be justified and integrated into AV programming under these ethical frameworks, each providing a distinct rationale for its implementation.

Harm Confinement and Lawfulness (HCL). The participants' choice in scenario 1 and 2 when ODS was removed, reflects a prioritisation of moral and legal principle of avoiding harm to innocent third parties and transferring harm to the responsible party that has caused the emergency situation. If harm is not limited to this extent, anyone could potentially be struck by an AV, requiring everyone to be constantly vigilant of this danger. Therefore, consideration for the legal implication of a decision made by the AV and to the extent that other road users have acted lawfully takes precedence over saving the greater number of people. The survey results indicate that Australians consider rule-breakers (such as jaywalkers) as more ethically liable to suffer harms, and they preferred measures that prevented death to innocents. This is indeed the very fact that appeared to be missing from previous studies, including the MM experiment and the infamous Trolley problem (Awad et al., 2018).

HCL makes harm distribution dynamic, as it will depend on scenario specific variables. The rule prioritising legal considerations and the lawful actions of road users over saving the greater number of people aligns best with deontological ethics. Deontological ethics focus on adherence to rules and duties, making it suitable for programming an AV to respect legal responsibilities even at the expense of maximising overall lives saved. Legal responsibility could be integrated into a decision tree, where the default actions taken by the AV change based on the legal implications of each scenario. In contrast, utilitarianism, which emphasises maximising overall happiness and minimising harm, would generally favour saving the greater number of people regardless of their legal status, making it incompatible with this rule. Similarly, Rawlsian ethics, which focuses on fairness and the protection of the least advantaged, does not inherently prioritise legal considerations, thus also making it a less suitable framework for this specific rule.

Harm Prevention and Prioritisation (HPP). Participants' preferences in scenario 4, when the ODS was removed, highlighted a strong consensus to prioritise preventing harm rather than intentionally causing another person's death, especially considering that all parties involved were part of the original scenario and at fault. Killing implies that you are directly respon-

sible for the death of another person and something that, for liability reasons, remains critically important to AV manufacturers. Allowing someone to die on the other hand, entails much less responsibility on your part, as there were some contributory factors already in motion that you did not initiate or otherwise control.

Given that both the pedestrian and the motorcyclist were at fault, a utilitarian AV would calculate the total expected harm from each option. If the AV determines that braking to avoid the pedestrian results in less overall harm (e.g., the motorcyclist may have a higher chance of survival even in a collision), it would choose this option. The AV could be programmed to use advanced sensors and algorithms to estimate the severity of injuries or likelihood of death, making an informed decision that minimises total harm. The deontological AV would be programmed to brake, as this action avoids the intentional harm to the pedestrian, despite the pedestrian's negligence. The AV would prioritise adhering to moral rules over the consequences, focusing on the duty to avoid causing harm. Similarly, a Rawlsian AV would be programmed to brake, prioritising the pedestrian who is directly in the path and at immediate risk. The AV would aim to distribute the risks more equitably, ensuring that harm is not disproportionately inflicted on the more vulnerable party, in this case, the pedestrian.

7 PROPOSED ETHICAL POLICY

An ethical policy consists of broad, abstract values that guide the development of socially acceptable AVs, ensuring they align with societal norms and expectations. The specific decision-making models and normative ethics can be too complex for the average user to understand, so an ethical policy is necessary to explain how AVs will make decisions (Liu et al., 2021). Based on my research findings, the proposed ethical policy includes the following four requirements, which are grounded in Australian moral preferences:

Injury Over Sacrifice (IOS). In any scenario where the AV can choose between causing injury or death, the AV will prioritise actions that minimise fatalities. Severe injuries are considered more acceptable than deaths.

Harm Confinement and Lawfulness (HCL). When all available options lead to death, the AV will aim to transfer harm to those who are responsible for creating the hazardous situation. The AV confines harm to those directly involved and ensures innocent bystanders are not harmed. The AV's actions will be guided by legal implications and the extent to which other road users have acted lawfully. Legal principles and moral responsibility take precedence over simply saving the greatest number of people.

Harm Prevention and Prioritisation (HPP). In situations where multiple parties are equally at fault and have all contributed to the dangerous situation, the AV will avoid taking direct actions that would intentionally cause any individual's death, opting instead to remain passive. This avoids intentional killing, reflecting legal and ethical priorities. If multiple principles apply to a moral dilemma, the AV's algorithm should evaluate the scenario based on all applicable principles. The above principles work together and are not standalone. The AV should dynamically balance IOS, HCL, and HPP, ensuring a holistic and ethically sound approach to decision-making.

Objective Decision System (ODS). When faced with scenarios where decision-making is hindered by technological limitations, time constraints, or other factors and the IOS, HCL or HPP principles are rendered ineffective, the AV must employ the ODS. This system will evaluate all viable and justified outcomes and select one at random, ensuring fairness, impartiality, and non-discrimination in the decision-making process.

7.1 Revisit Crash Scenarios

In this section I revisit the crash scenarios from the survey questionnaire and discuss how an AV programmed with the proposed ethical settings would handle each situation.

Scenario 1 - Brakes Failure on Mountain Road
Moral Dilemma: The AV must choose between swerving right to hit an innocent bystander or continuing straight off a cliff, resulting in the death of the passenger.

Principle Application:

- IOS: Not applicable as both outcomes lead to death.
- HCL: The bystander is not culpable for the emergency situation or the AV's mechanical failure. The AV should prioritise confining harm and not introducing new parties to the risk.
- HPP: The AV should avoid actively causing the death of an innocent bystander.
- ODS: If no clear decision can be made, the AV would randomly choose from justified options.

AV Response: The AV would continue straight off the cliff, resulting in the death of the passenger. This action aligns with HCL and HPP by avoiding harm to an innocent bystander who did not contribute to the dangerous situation.

Scenario 2 - Intersection with Illegally Crossing Pedestrians
Moral Dilemma: The AV can either swerve left to save five illegally crossing pedestrians, causing the death of a cyclist, or continue straight, potentially hitting one or more pedestrians.

Principle Application:

- IOS: Not applicable as all possible outcomes lead to death.
- HCL: The illegally crossing pedestrians are at fault. The AV should prioritise confining harm to the responsible parties and transferring harm to those who contributed to the dangerous situation.
- HPP: The AV should avoid actively killing the cyclist who is not at fault.
- ODS: If no clear decision can be made, the AV would randomly choose from justified options.

AV Response: The AV would continue straight, potentially hitting one or more of the pedestrians. This action aligns with HCL by confining harm to those who are responsible for the situation (the illegally crossing pedestrians) and avoids actively killing the cyclist, who is not at fault.

Scenario 3 - Highway Dilemma with Fallen Pallet
Moral Dilemma: The AV must choose between continuing straight and killing the passenger, swerving right and killing a helmetless motorcyclist, or swerving left and causing severe injury to a fully geared motorcyclist.

Principle Application:

- IOS: Applicable; the AV should avoid causing death by choosing an outcome that minimises fatalities.
- HCL: The helmetless motorcyclist is partially at fault for not wearing safety gear.
- HPP: The AV should avoid actively killing and prefer causing severe injury over death.
- ODS: If no clear decision can be made, the AV would randomly choose from justified options.

AV Response: The AV would swerve left, causing severe injury but survival of the fully geared motorcyclist. This action aligns with IOS by preventing death and with HPP by avoiding active killing. The helmetless motorcyclist, while partially at fault, is spared to prevent a fatality.

Scenario 4 - Pedestrian vs. Motorcyclist
Moral Dilemma: The AV must choose between braking to save a pedestrian, causing a motorcyclist to crash into the AV and die, or continuing and striking the pedestrian, saving the motorcyclist.

Principle Application:

- IOS: Not applicable as both outcomes lead to death.
- HCL: Both the pedestrian and the motorcyclist are considered negligent.
- HPP: The AV should avoid actively causing the death of either party.
- ODS: If no clear decision can be made, the AV would randomly choose from justified options.

AV Response: The AV would brake to save the pedestrian, causing the motorcyclist to collide with the AV and be killed. This action aligns with HPP by avoiding the active killing of the pedestrian, who is directly in the AV's path.

Scenario 5 - Intersection with Legally Crossing Pedestrians
Moral Dilemma: The AV must decide between swerving into a wall, risking the lives of five passengers, or proceeding straight, potentially endangering legally crossing pedestrians.

Principle Application:

- IOS: Not applicable as both outcomes lead to death.
- HCL: The pedestrians are legally crossing, while the AV's passengers are not at fault.
- HPP: The AV should avoid actively causing the death of the pedestrians who are acting lawfully.
- ODS: If no clear decision can be made, the AV would randomly choose from justified options.

AV Response: The AV would swerve into the wall, risking the lives of its five passengers to save the pedestrians. This action aligns with HCL by protecting the legally crossing pedestrians and with HPP by avoiding active killing.

Scenario 6 - Intersection with Legal Pedestrian Crossing
Moral Dilemma: The AV must choose between swerving into a tree, killing three passengers, or continuing straight and hitting a pedestrian legally crossing.

Principle Application:

- IOS: Not applicable as both outcomes lead to death.

- HCL: The pedestrian is legally crossing, while the passengers are not at fault.
- HPP: The AV should avoid actively causing the death of the legally crossing pedestrian.
- ODS: If no clear decision can be made, the AV would randomly choose from justified options.

AV Response: The AV would swerve to the right, colliding with a tree and resulting in the deaths of three passengers. This action aligns with HCL by protecting the legally crossing pedestrian and with HPP by avoiding active killing.

8 CONCLUSION

The research set out to explore how a publicly accepted ethical setting for an AV could look like. It presented different ethical theories, decision-making models and moral variables that could help solve moral dilemmas faced by AVs. The objective was to evaluate respondents' perceptions and expectations regarding personal AVs in relation to different ethical settings. The survey questions were designed to eliminate any biased or redundant criteria based on previous research, ensuring that participants' decisions were informed by both ethical and legal considerations through the inclusion of legal implications for each crash scenario.

A significant aspect of the research was the introduction of a randomised choice option, referred to as the ODS. In scenarios where the AV encounters a moral dilemma, such as the decision between braking to avoid hitting a pedestrian or not braking, resulting in harm to the motorcyclist behind, the ODS allows the AV to evaluate all viable outcomes based on a moral framework and then select one at random. The findings revealed that both utilitarianism and Rawlsian ethics were unsuitable for aligning with the participants' preferences, as the core principles of these ethical theories do not adequately account for the different roles of participants in harm distribution. Consequently, the proposed ethical setting is grounded in deontology, utilising a rule-based decision-making model with moral variables, including IOS, HCL, HPP, and ODS as redundancy. This shift represents a departure from the utilitarian harm minimisation approach observed in other studies.

Research questionnaires, like the one I conducted, are valuable for simplifying complex real-world issues, focusing on core ethical considerations (Pojman, 1990). They can help test initial programming ideas for AVs, as algorithms that contradict societal moral expectations are likely to hinder the smooth adoption of AVs. Furthermore, ethical decision-making cannot rest solely with engineers or ethicists; a consensus is essential, as solutions to programming moral dilemmas in AVs should not deter potential buyers. Engaging the public can spark curiosity about the various ways AVs could be programmed to address ethical dilemmas (Nyholm, 2018).

It has been observed that respondents' decisions shift with the level of concreteness in experiments (Francis et al., 2017). For future research, I recommend using virtual reality to create realistic crash scenarios, allowing participants to engage with different dilemmas and choose outcomes they deem most acceptable. Additionally, I have outlined the advantages of the ODS in AV programming. Despite initial concerns about using lotteries for determining legal outcomes, ODS should not be dismissed in AV crash algorithms. Decision-making by lot is often simple, objective, and inexpensive, and my research suggests that it is both understandable and acceptable to many individuals. Thus, I believe that lotteries, more than other decision-making devices, merit serious consideration by future researchers, AV manufacturers, and policymakers.

Developing ethical autonomous machines is one of the most challenging topics in artificial intelligence today. It extends beyond resolving ethical dilemmas; setting realistic expectations with users and the public is crucial for market acceptance and adoption. These challenges are common across automotive manufacturers engaged in this emerging field, not limited to specific companies. AVs promise substantial benefits while also presenting unintended effects that are difficult to predict, and the technology is progressing rapidly, regardless of ethical considerations (Lin, 2016). This juncture represents an ideal opportunity to discuss the ethical values and principles best suited for self-driving cars, laying the groundwork for future improvements.

REFERENCES

Australian Bureau of Statistics (2021). Australia migration 2019-20 financial year. https://www.abs.gov.au/statistics/people/population. Accessed: 2021-06-20.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., and Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729):59–64.

Bonnefon, J.-F., Shariff, A., and Rahwan, I. (2015). Autonomous vehicles need experimental ethics: are we ready for utilitarian cars. *arXiv preprint arXiv:1510.03346*.

Bonnefon, J.-F., Shariff, A., and Rahwan, I. (2016). The

social dilemma of autonomous vehicles. *Science*, 352(6293):1573–1576.

Botes, A. (1996). Understanding social research. *Health SA Gesondheid : Journal of Interdisciplinary Health Sciences*, 1.

Campbell, M., Egerstedt, M., How, J. P., and Murray, R. M. (2010). Autonomous driving in urban environments: approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4649–4672.

Coca-Vila, I. (2018). Self-driving cars in dilemmatic situations: An approach based on the theory of justification in criminal law. *Criminal Law and Philosophy*, 12(1):59–82.

Curtis, P. and Hedlund, J. (2005). Reducing deer-vehicle crashes. report funded by the insurance institute for highway safety. cornell university, ithaca.

Duxbury, N. (1999). *Random Justice: On Lotteries and Legal Decision-Making*. Oxford University Press UK.

Dworkin, R. (2011). *Justice for hedgehogs*. Harvard University Press.

Etienne, H. (2020). When ai ethics goes astray: A case study of autonomous vehicles. *Social Science Computer Review*, page 0894439320906508.

Farhadi, A., Endres, I., Hoiem, D., and Forsyth, D. (2009). Describing objects by their attributes. In *2009 IEEE conference on computer vision and pattern recognition*, pages 1778–1785. IEEE.

Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., and Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific Reports*, 7(1):1–11.

Gomez, L. R. P., Szybalski, A. T., Thrun, S., Nemec, P., and Urmson, C. P. (2014). Transportation-aware physical advertising conversions. US Patent 8,630,897.

IEEE (2016). Ethically aligned design: A vision for prioritizing wellbeing with artificial intelligence and autonomous systems.

Lin, P. (2013). The ethics of autonomous cars. the atlantic, october 8.

Lin, P. (2016). Why ethics matters for autonomous cars. In *Autonomous driving*, pages 69–85. Springer, Berlin, Heidelberg.

Liu, Y., Sun, L., Wang, Y., and Ren, Y. (2021). Vision-based obstacle detection and recognition for autonomous vehicles. *Measurement*, 184:109984.

Luetge, C. (2017). The german ethics code for automated and connected driving. *Philosophy & Technology*, 30(4):547–558.

Noothigattu, R., Gaikwad, S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., and Procaccia, A. (2018). A voting-based system for ethical decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap, i. *Philosophy Compass*, 13(7):e12507.

Pojman, L. (1990). Discovering right and wrong. *Belmont, CA: Wadsworth*.

Rachels, J. and Rachels, S. (2012). *The Elements of Moral Philosophy 7e*. McGraw Hill.

Rafiee, A., Wu, Y., and Sattar, A. (2023). Philosophical and legal approach to moral settings in autonomous vehicles: An evaluation. In *Social Licence and Ethical Practice*, volume 27, pages 95–114. Emerald Publishing Limited.

Rescher, N. (1960). Choice without preference. a study of the history and of the logic of the problem of "buridan's ass". *Kant-Studien*, 51(1-4):142–175.

Shadrin, S. S. and Ivanova, A. A. (2019). Analytical review of standard sae j3016 «taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles» with latest updates. *Avtomobil'. Doroga. Infrastruktura.*, (3 (21)):10.

Shahriari, K. and Shahriari, M. (2017). Ieee standard review—ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, pages 197–201. IEEE.

Zimmermann, T. (2014). Justification or excuse? the collision of positive duties in german criminal law. *The Journal of Criminal Law*, 78(3):263–276.