



**UNIVERSITY OF PADOVA**

**Department of General Psychology**

**Bachelor's Degree Course in Psychological Science**

**Final Dissertation**

Risk perception in traditional and autonomous driving scenarios

**Supervisor**

**Professor Ph.D** Andrea Spoto

**Candidate:** Armağan Fırat Özçelik

**Student ID number:** 2041177

Academic Year 2023 - 2024

## **Table of Contents**

<b>Abstract</b> .....	3
<b>Chapter 1: Introduction</b> .....	4
—1.1 Introducing Morality .....	4
— 1.2 Autonomous Driving and the “AV Dilemma” .....	5
— 1.3 Moral Foundation Theory .....	8
<b>Chapter 2: Research</b> .....	10
— 2.1 Objective of the Study .....	10
— 2.2 Participants.....	12
— 2.3 Experimental Procedure .....	12
—2.4 Experimental Material.....	13
—2.5 Data Analysis .....	15
<b>Chapter 3: Results</b> .....	16
<b>Chapter 4: Conclusion</b> .....	21
<b>References</b> .....	24

## ABSTRACT

One important turning point in the development of transportation and technology is the progress made in autonomous driving. As we will integrate these machines into our lives in the next future, the interest in the moral evaluation of their decisional algorithms has substantially increased (Evans et al, 2020). Morality in Autonomous Vehicles (AV) is a tangled novel topic which is still covered by traditional moral standards. This technology has the burden of requiring complex decisions, some of which might also include unavoidable harm (Bonnefon et al., 2016), a condition in which at least one party is negatively affected due to accidents or manufacture errors. Although it may seem improbable, unlike humans they have to be programmed beforehand on the basis of behavioral rules and social norms (Faulhaber et al. 2018). Potential scenarios include deciding between advantages for the self, other passengers or third parties, all situations that might force the AV to make a decision between two or more unpleasant and unsatisfactory choices. The fact that moral decisions have to be predetermined opens up the question of whether they should protect passengers or other parties, such as the AV owner or the majority of the characters involved.

## 1. INTRODUCTION

### 1.1 Introducing morality

Morality is a fundamental aspect of human psychology, influencing emotions, decision-making, and social behavior. It encompasses how people integrate moral aspects on reasoning and judgments, the underlying psychological mechanisms, and how the results of these processes influence behavior (Haidt, 2012). Additionally, the field of moral psychology explores how moral beliefs and behaviors are influenced by social contexts and interpersonal relationships. Research indicates that moral reasoning is not only a cognitive process but also a social one, influenced by factors such as group dynamics, cultural norms, and personal experiences (Gibbs, 2003).

In psychological research, morality is often explored to understand how individuals make ethical decisions and how these decisions are shaped by different moral paradigms. This research line also intersects the role of emotions, social influences, and cognitive processes in moral decision-making. For instance, Jonathan Haidt's work on moral psychology shows how moral judgments are influenced by emotional responses and social context, rather than purely rational deliberation (Haidt, 2012) Two major moral theories that are frequently studied in this context of research are Utilitarianism and Deontology. Utilitarianism is a theory that states the morality of an action is determined by its outcome. The core principle of utilitarianism is to maximize overall happiness or utility. According to this view, an action is morally right if it produces the greatest amount of good for the greatest number of people. Utilitarianism can be divided into "act utilitarianism" and "rule utilitarianism" "Act utilitarianism" evaluates each action

individually, assessing whether it maximizes happiness. “Rule utilitarianism”, on the other hand, focuses on adhering to rules that, in general, lead to the greatest good. This distinction is crucial as act utilitarianism can sometimes justify morally questionable actions if they result in a net positive outcome, whereas rule utilitarianism aims to establish consistent ethical guidelines that promote overall welfare (Harsanyi, 1977). The application of utilitarian principles in psychological studies often involves examining how people weigh the consequences of their actions and the extent to which they consider the well-being of others.

Deontologism, or deontological ethics, is a theory that focuses on the innate morality of actions rather than their consequences. According to deontological ethics, certain actions are morally obligatory, permissible, or forbidden based on a set of rules or duties. This theory originates from the philosophy of Immanuel Kant, who argued that moral principles are grounded in rationality and should be followed unconditionally (Kant, 1785). Unlike utilitarianism, deontologism asserts that some actions are intrinsically right or wrong, regardless of their outcomes.

## 1.2 Autonomous Driving and the “AV dilemma”

Autonomous vehicles (AVs) are means of transportation designed to navigate and maneuver with reduced or any form of human intervention. There are five levels of vehicle automation that provides an understanding of the progressive stages towards fully autonomous vehicles. According to the National Highway Traffic Safety Administration (NHTSA), these levels range from no automation (Level 0) to full automation (Level 5). At Level 0, the driver is entirely responsible for all vehicle operations, including steering, braking, and acceleration. Level 1 introduces specific control functions that assist the

driver, such as adaptive cruise control or automated parking systems, but the driver remains engaged with the driving task at all times. Level 2 automation combines at least two automated functions working in unison to relieve the driver of controlling those specific functions, like simultaneous use of adaptive cruise control and lane centering. Level 3 automation allows the driver to cede full control of all critical driving tasks under certain conditions, although the driver must be prepared to intervene if necessary. In Level 4, the vehicle is capable of performing all driving tasks and monitoring the environment without any driver intervention, but only in certain conditions or environments. The ultimate goal, Level 5 automation, envisions vehicles that are completely autonomous, requiring no human intervention regardless of the conditions (Davidson & Spinoulas, 2015). These vehicles must be programmed to make split-second decisions that sometimes may also have ethical implications, such as prioritizing the safety of AV's passengers or the protection of a certain number of pedestrians facing an unavoidable accident scenario (Lin, 2016; Gurney, 2017). The development and implementation of AVs require a high-level of complexity that also converges towards ethical assumptions and social norms. The moral dilemmas associated with autonomous driving (i.e., the AV dilemma) is a typical example of this concept. This involves integrating various moral theories and frameworks that guides the decision-making processes of AVs (Goodall, 2014). For instance, the Trolley Problem, a classic moral dilemma, which is a well-known ethical dilemma firstly introduced by Philippa Foot (1967) and later fully fledged out by Judith Jarvis Thomson in the form of the "Footbridge Dilemma" (1976) It presents a scenario in which a trolley is headed towards five people tied to a track. The brakes have failed, and the only way to save the five people is to divert the trolley onto another track, where it will kill one person instead. The central question of the Trolley Problem is

whether it is morally permissible to turn the trolley, sacrificing one life to save five (Foot, 1967) it has been adapted to examine how they should respond in more “realistic” scenarios where harm is unavoidable (Bonnefon et al, 2016). Resembling the traditional trolley problem, an AV prioritize the safety of its passengers over a higher number of pedestrians, or should it follow a utilitarian approach, minimizing overall harm regardless of who is affected? These questions are central in the discussion on the ethical programming of AVs and have significant implications for their acceptance and deployment in society.

Autonomous vehicles (AVs) present significant moral dilemmas that necessitate a psychological perspective to understand how ethical decision-making can be programmed into these systems. The Moral Machine experiment by Bonnefon, Shariff, and Rahwan (2016) revealed that people generally prefer AVs that follow a utilitarian approach, making decisions that minimize overall harm. This preference aligns with the utilitarian principle, the greatest good for the greatest number, which is essential for AVs to make ethical choices in unavoidable harm scenarios. However, the psychological complexity of morality becomes evident when considering the reluctance of individuals to use AVs programmed to sacrifice their passengers for the greater good. For instance, in a situation where an AV must choose between hitting a group of pedestrians or swerving to hit a single pedestrian, the utilitarian approach would dictate minimizing the total harm, even if it means sacrificing one individual to save many, highlighting a discrepancy between abstract moral principles and personal safety concerns, as individuals may theoretically support utilitarian ethics but are less willing to apply them when their own safety or that of their loved ones is at risk.

Shah et al. (2022) dives further into the psychological reasonings influencing moral judgments in AV scenarios. The research indicates that people's decisions are pragmatically bounded, meaning moral preferences can change and evolve based on contextual factors and emotional connections. For example, individuals might prioritize saving loved ones over strangers or deny responsibility when multiple stakeholders are involved. These psychological insights are crucial for developing AV algorithms that can adapt to varying moral contexts and societal expectations, ensuring that AVs operate in ways that resonate with human moral intuitions. This adaptability is particularly important in situations where AVs must make split-second decisions. The role of individual experience and contextual factors becomes evident in how people endorse different moral outcomes in dilemmas. These factors heavily influence the likelihood of choosing one moral path over another, as personal history, cultural background, and ingrained societal values shape moral intuitions. This connection between individual experiences and normative influences makes understanding the broader moral frameworks that guide decision-making, such as those explored in the Moral Foundations Theory a crucial step.

### 1.3 Moral Foundation Theory

Moral Foundations Theory (MFT), which posits that there are several modular foundations of morality that have evolved to solve specific challenges (Graham et al., 2009). Moral Foundations Theory (MFT) offers a framework for understanding moral reasoning by identifying several fundamental moral domains, or "foundations," that guide human moral thinking. MFT suggests that moral reasoning extends beyond the traditional focus on fairness and harm, to several different domains.



According to Graham et al. (2013), these foundations provide a basis for understanding the intuitive and emotional aspects of moral cognition.

- The loyalty/betrayal dimension measures the significance of loyalty to one's group, family, or nation, reflecting values associated with solidarity, patriotism, and group cohesion, and the emotional responses to betrayal and disloyalty.
- The authority/subversion dimension gauges respect for tradition, authority, and social order, including attitudes towards obedience to rules, respect for leadership, and maintenance of societal hierarchies.
- The sanctity/degradation dimension evaluates concerns about purity, sanctity, and contamination of body and soul, associated with moral emotions like disgust and reverence, often pertaining to religious beliefs, dietary practices, and attitudes towards sexuality.
- The care/harm dimension is focused on protecting others from harm and promoting compassion and kindness. It pushes individuals to prevent suffering and show empathy, with a strong concern for the well-being and safety of others.
- The fairness/cheating dimension is centered around justice, equality, and reciprocity. It motivates individuals to act in ways that promote fairness and to oppose cheating or dishonesty. A strong emphasis on fairness leads to a desire for equal treatment and a sense of moral outrage when others are treated unfairly or when deceitful behavior occurs (Graham et al., 2013).
- The liberty/oppression dimension is focused on personal freedom and autonomy, resisting domination and coercion by others. It reflects a strong sense of individual rights and is particularly concerned with fighting against oppression or authoritarian control. People who value liberty are sensitive to any infringement

on personal freedom and advocate for protecting individual rights and autonomy (Iyer et al., 2012).

These sub-scales collectively provide a comprehensive understanding of an individual's moral profile, indicating which foundations are most influential in their moral reasoning and decision-making processes. Research using these sub-scales has shown that different groups prioritize these foundations differently, with conservatives typically scoring higher on loyalty, authority, and sanctity, and liberals scoring higher on care and fairness (Graham et al., 2013). Later, Maxwell and Narvaez (2013) expand on the impact of MFT in the field of moral psychology, their research argues that MFT's broader perspective on morality, has reshaped the understanding of moral development and education. By recognizing multiple moral domains, MFT facilitates a more comprehensive approach to studying moral judgments and behaviors, accounting for the variability observed across different cultural and individual contexts. MFT's foundations are investigated through the MFQ, a widely utilized instrument designed to measure individuals' support for various moral foundations. The questionnaire will be presented in detail in the Experimental Materials section.

## 2. THE RESEARCH

### 2.1 Objective and outline of the present dissertation

Building on the foundational literature in moral psychology, presented in the introduction, this study examines how individuals' moral foundations, as measured by the Moral Foundations Questionnaire (MFQ), might influence their decision-making in the context of autonomous vehicles (AVs). This research hypothesizes that these moral foundations

will significantly impact the likelihood of endorsing utilitarian outcomes in AV-related moral dilemmas.

Specifically, it is anticipated that individuals who score higher on the care (MFQ-H) and fairness (MFQ-F) dimensions will be less inclined to support utilitarian decisions. This expectation arises from the emphasis these foundations place on the protection of individuals and equitable treatment, which may conflict with the utilitarian principle of maximizing overall welfare. The complexities involved in aligning these moral intuitions with utilitarian decisions in AV scenarios underscore the importance of understanding these relationships (Awad et al., 2018).

Conversely, individuals with higher scores in the authority (MFQ-A) and loyalty (MFQ-I) dimensions are hypothesized to be more likely to endorse utilitarian outcomes, reflecting a greater emphasis on maintaining social order and group cohesion, which may align with the utilitarian approach to decision-making in AV scenarios (Awad et al., 2018).

The results chapter will present descriptive scores of the selected sample from a reference population of university students. Correlations between the moral foundations dimensions and with levels of acceptance of the utilitarian decision will be described, and the potential role of MFQ's dimension on the likelihood of endorsing the utilitarian AV behavior will be investigated with the use of a Generalized Mixed Effects Linear Model. Finally, the discussion will interpret these results in the context of our hypothesis, discussing some potential implications for the ethical programming of AVs and identifying areas for future research.

## 2.2 Participants

The present dissertation presents evidence from a subset of 61 participants, extracted from the original dataset, which was composed of 206 participants. This subset corresponds to a specific experimental condition of the original experiment, involving three conditions in a between-subject experimental design. All participants were Italian native speakers. The demographic data collected includes age, gender, and years of education.

The participants' ages ranged from a minimum of 18 to a maximum of 60, producing an age range of 42 years. Gender distribution among the participants is as follows:

Males: 44.26% Females: 55.74%. with no participants identifying as non-binary or providing an unspecified gender.

Regarding education, participants reported their years of education rather than the highest degree obtained. The reported years of education ranged from a minimum of 2 years to a maximum of 24 years. The overall mean years of education among participants was 16.607, with a standard deviation of 3.018. When the data is split by gender, no significant difference is observed: males reported an average of 16.593 years of education, slightly below the general mean, and females reported an average of 16.618 years, slightly above the general mean.

## 2.3 Experimental Procedure

The study was designed and administered using Qualtrics software. The study link was shared through social networks and institutional communication channels targeting Italian

university students. Participants were instructed to complete the survey exclusively on laptops, avoiding smartphones or tablets. Initially, participants were required to respond to three dilemmas, which were presented in a randomized order, followed by the Moral Foundations Questionnaire (MFQ).

## 2.4 Experimental Materials

### *2.4.1 The Moral foundation Questionnaire.*

The 32-item Moral Foundations Questionnaire (MFQ), as detailed by Graham et al. (2009), was employed to assess participants' endorsement of the established moral foundations: care, fairness, authority, loyalty, and purity. Each moral foundation is evaluated through six items rated on a six-point Likert scale which ranges from "Strongly Disagree" to "Strongly Agree." Each set of six items corresponding to a specific moral foundation—care, fairness, authority, loyalty, and purity—is averaged to calculate the overall score for that foundation. Average scores for each foundation were calculated using the syntax provided by the MFQ authors. Additionally, the MFQ was supplemented with nine items from Iyer et al. (2012) to assess endorsement of the liberty foundation. These liberty items were randomly placed within their respective section of the MFQ to reduce response bias and were further randomized for each participant.

### 2.4.2 The AV dilemma

Provided below is an example of a Moral dilemma in an AV scenario

*“You are on board of your autonomous driving vehicle, cruising on a main road respecting all speed limits. Suddenly, five pedestrians proceed to cross the road, on your same trajectory. They are now in front of the automobile, which, although working properly and having done nothing wrong, now lacks the time to brake safely.*

*The vehicle now has two choices:*

- *To proceed straight, protecting you but running over the five pedestrians, who will certainly die.*
- *To steer suddenly to your left, towards a roadblock. The five pedestrians will be safe, but you will lose your life in the crash.”*



Figure 1: For representative purposes, the figure describes a graphical representation of the moral dilemma: the subject (inside the red car) is about to impact the pedestrians (black silhouettes), unless the autonomous vehicles decides to steer towards the roadblock.

## 2.5 Data analysis

The analysis begins with the presentation of descriptive statistics for the MFQ sub-scales (Authority, Fairness, Purity, Loyalty, and Care) and other continuous variables, including utilitarian acceptability. Following the descriptive statistics, the results section will present correlation analyses between the MFQ sub-scales and variables such as utilitarian acceptability. Pearson's correlation coefficients are used to explore the relationships between these variables, helping to identify potential associations that may inform further analysis. The correlation matrix is provided to give a comprehensive view of how these moral foundations relate to each other and to the outcome variables.

The final part of the analysis involves the application of a Generalized Linear Mixed Effects Model (GLMER). This model is used to assess the impact of the MFQ sub-scales on participants' responses to moral dilemmas, with a specific focus on utilitarian acceptability. The GLMER accommodates both fixed effects (such as the MFQ sub-scales) and random effects (such as individual differences), allowing for a more nuanced

exploration of how these moral foundations influence ethical decision-making in the context of autonomous vehicles.

### 3. RESULTS

#### 3.1 Descriptives

To gain an initial understanding of the data, descriptive statistics were calculated for the Moral Foundations Questionnaire (MFQ) sub-scales—Authority (MFQ-A), Care (MFQ-H), Fairness (MFQ-F), Loyalty (MFQ-I), and Purity (MFQ-P)—as well as for variables related to utilitarian acceptability and plausibility in the context of autonomous vehicles (AVs). All of the statistics can be found in Table 1

Table 1 - Statistics of MFQ sub-scales

Variable	Mode	Median	Mean	Std. Deviation	Min	Max
MFQ-A Total	17.000	17.000	17.689	4.689	7.000	29.000
MFQ-H Total	27.000	27.000	26.410	5.074	7.000	35.000
MFQ-F Total	30.000	29.000	28.000	4.590	7.000	36.000
MFQ-I Total	18.000	19.000	19.639	5.348	9.000	30.000
MFQ-P Total	11.000	15.000	15.639	4.768	7.000	31.000



### 3.2 Correlations

To analyze the correlation between the respective sub-scales, the acceptability of morally accepting utilitarian and the plausibility of utilitarian reasoning in the cases of AV's in further detail, a correlation analysis was computed which is shown in Table 2.

*Table 2 Correlation Results*

Variable	moral_acc_utilitarian_AV	plaus_util_AV	MFQ A TOTAL	MFQ F TOTAL	MFQ I TOTAL	MFQ P TOTAL	MFQ H TOTAL
moral_acc_utilitarian_AV	-	0.414* **	0.142	0.225	0.121	0.147	0.192
plaus_util_AV	0.414***	-	0.252*	0.129	0.105	0.126	0.68
MFQ A TOTAL	0.142	0.252*	-	0.283*	0.564** *	0.647** *	0.144
MFQ F TOTAL	0.225	0.129	0.283*	-	0.439** *	0.225	0.691** *
MFQ I TOTAL	0.121	0.105	0.564** *	0.439** *	-	0.469** *	0.375**
MFQ P TOTAL	0.147	0.126	0.647** *	0.225	0.469** *	-	0.176
MFQ H TOTAL	0.192	0.068	0.144	0.691** *	0.375**	0.176	-

Pearson's correlations were computed between moral acceptability of utilitarian judgments, the perceived plausibility of utilitarian arguments, and the total scores of various Moral Foundations Questionnaire (MFQ) sub-scales: Authority, Fairness, Loyalty,

Purity, and Harm. The statistical significance of these correlations is indicated, with asterisks representing significance levels (\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ ).

When examining the relationship between MFQ sub-scales and moral acceptability of utilitarianism, the Authority ( $r = .142$ ), Fairness ( $r = .225$ ), and Purity ( $r = .147$ ) foundations showed positive, but non-significant correlations when it came to moral acceptance. This implies a weak or negligible connection between these moral foundations and participants' acceptability of utilitarian judgments. Similarly, the Loyalty ( $r = .121$ ) and Harm ( $r = .192$ ) foundations demonstrated non-significant relationships with the utilitarian moral acceptance.

In terms of the perceived plausibility of utilitarian arguments, a significant positive correlation was found with the Authority sub-scale ( $r = .252$ ,  $p < .05$ ). This suggests that individuals who place greater emphasis on authority-based moral reasoning are more likely to perceive utilitarian arguments as plausible. Other moral foundations, such as Fairness, Loyalty, Purity, and Harm, did not show any significant relationships with the perceived plausibility of utilitarian arguments.

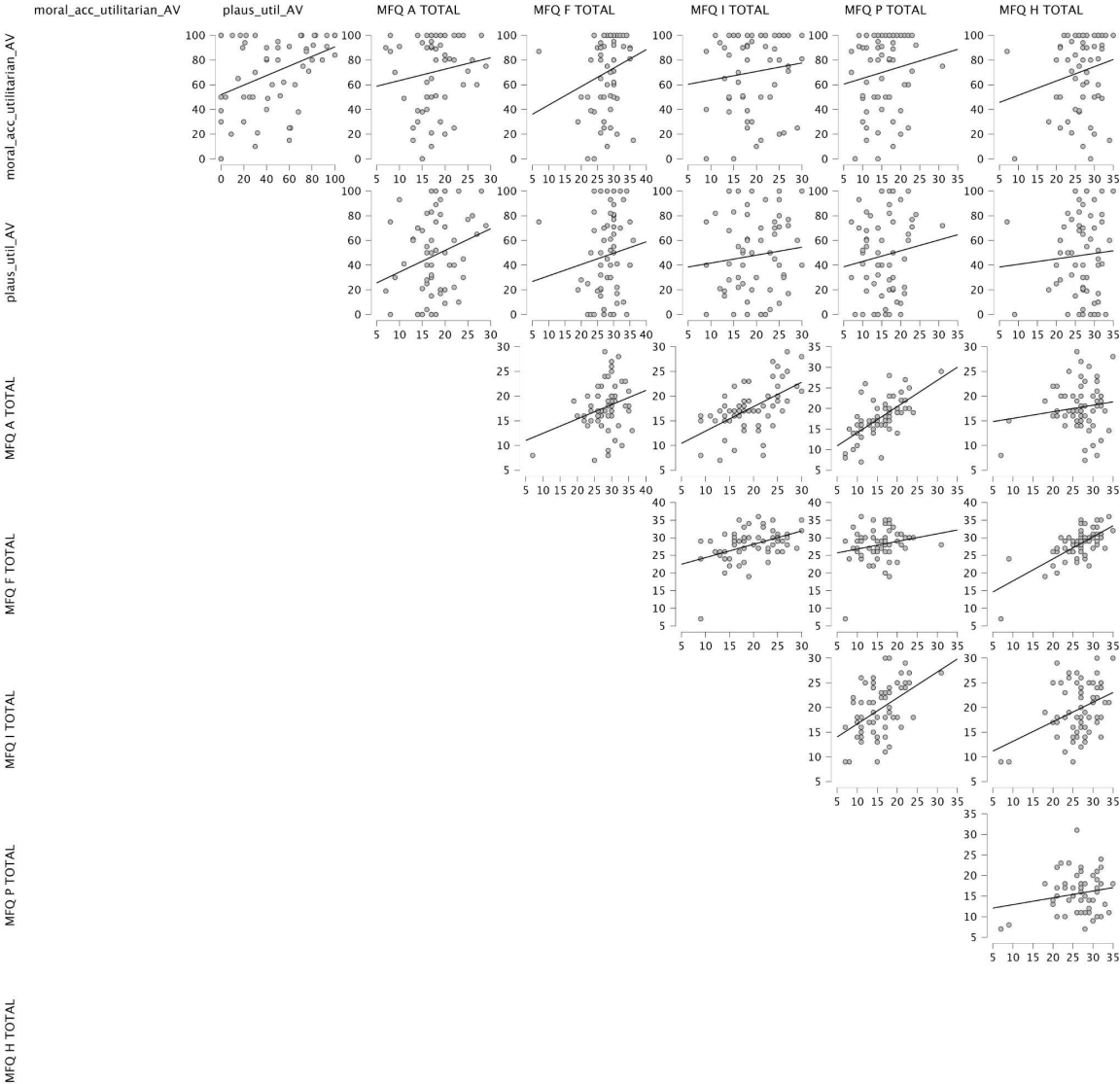
Correlations described in the previous matrix are now graphically represented by the use of scatter plots, which show several relationships between moral acceptance of utilitarian judgments, the plausibility of utilitarian arguments, and the sub-scales of the Moral Foundations Questionnaire (MFQ).

The scatter plot also shows weaker relationships between the moral acceptance of utilitarianism and the various MFQ sub-scales. For instance, the plot between moral acceptance of utilitarianism and the Authority sub-scale shows a slight upward trend,

suggesting a weak positive relationship. However, other sub-scales, such as Fairness, Harm, and Loyalty, do not show strong patterns in their relationship with moral acceptance of utilitarianism, as evidenced by the relatively flat trends in the scatter plots.

In contrast, the relationships between the sub-scales themselves show stronger correlations. For example, the scatter plots between Authority and Loyalty, and Loyalty and Purity, exhibit stronger positive trends, suggesting that these moral foundations are more closely aligned with one another. The detailed plots can be found in table 3.

Table 3 - Correlation Plots



### 3.3 Generalized mixed effects linear model

Three moral dilemmas (rD1, rD2, and rD3) were analyzed, each presenting binary response options (0 = morally acceptable, 1 = morally unacceptable).

The fixed effects estimates (Table 4) provide insights into the specific effects of each MFQ sub-scale on moral decisions. Any significant effect of MFQ sub-scales on the likelihood of selection the utilitarian as an option was observed. Descriptively, a tendency to observe a higher probability of endorsing the utilitarian option can be detected for higher scores of the fairness sub-scale, but the effect did not reach statistical significance.

Table 4 - Fixed effects estimates

Term	Estimate	SE	t	p
Intercept	-3.926	2.991	-1.313	0.189
MFQ A TOTAL	0.025	0.135	0.185	0.854
MFQ H TOTAL	0.005	0.128	-0.041	0.967
MFQ F TOTAL	0.247	0.150	1.644	0.100
MFQ I TOTAL	-0.076	0.111	-0.685	0.494
MFQ P TOTAL	-0.095	0.127	-0.747	0.455

#### 4. CONCLUSION

This dissertation's goal was to explore the relationship between moral foundations and ethical decision-making in autonomous vehicle (AV) scenarios, specifically focusing on whether specific moral foundations influence the likelihood of selecting a utilitarian outcome. The study employed the Moral Foundations Questionnaire (MFQ) to assess these moral dimensions and examined how they correlate with participants' responses to moral dilemmas presented in the context of AV decision-making.

The results indicate that the hypothesized relationships between moral foundations and the endorsement of utilitarian outcomes in autonomous vehicle (AV) moral dilemmas were not fully supported by the data. Specifically, while existing literature suggests that individuals with higher scores in the care and fairness dimensions of the Moral Foundations Questionnaire (MFQ) are generally less likely to support utilitarian decisions (Haidt, 2012), this study did not observe a significant negative correlation between these dimensions and the likelihood of endorsing utilitarian outcomes.

Prior research has shown that individuals scoring higher on the authority and loyalty dimensions might be more inclined to endorse utilitarian decisions.(Awad et al., 2018; Graham et al., 2013). However, the anticipated positive correlations between these dimensions and utilitarian decision-making were also not observed to the expected degree.

This study has significant implications for the development of ethical frameworks in the context of autonomous vehicles (AVs). The observed limited influence of moral foundations such as fairness and purity on utilitarian decision-making raises critical questions about the design and programming of AVs to ensure they make ethical decisions that align with diverse human moral intuitions.

Traditional moral theories like utilitarianism, have often been proposed as foundational frameworks for programming ethical decisions in AVs (Bonneton, et al., 2016). However, this study's results suggest that relying exclusively on utilitarian principles may overlook important moral dimensions that resonate with human values.

The findings also point out that ethical frameworks should be more inclusive of various moral perspectives to better align with human moral intuitions. Integrating diverse ethical theories, including deontological and virtue ethics perspectives, could lead to more comprehensive ethical decision-making models. By incorporating principles beyond utilitarianism, such as respect for individual rights or the cultivation of moral virtues, AVs could be better equipped to handle the complex moral landscapes they will inevitably encounter.

However, this study has limitations that should be acknowledged. The sample size, although sufficient for a first statistical investigation on potential trends between outcomes and predictors, may not capture the full diversity of moral perspectives present in a larger population. Additionally, the use of hypothetical text-based scenarios, while useful for an experimental investigation that aims to control confounding variables, may not fully replicate the complexities of real-world decision-making in AV contexts.

Future research should explore these moral foundations further, perhaps with larger and more diverse samples, to better understand how different moral intuitions interact in the context of AV decision-making. Additionally, exploring other ethical frameworks beyond utilitarianism and deontology could provide a more comprehensive understanding of how to program AVs to align with human moral values. Research could also benefit from examining real-world data on AV decisions to see how these moral theories play out in practice.

In conclusion, while this study has provided valuable insights into the relationship between moral foundations and ethical decision-making in AVs, it also highlights the need for a more nuanced approach to understanding and integrating human morality into autonomous systems. The findings challenge the simplicity of applying traditional ethical theories to complex, real-world scenarios and call for ongoing research into how AVs can be ethically programmed in a way that aligns with the diverse moral values of society.

## References

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*, 563(7729), 59-64. <https://doi.org/10.1038/s41586-018-0637-6>
- Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Nature Human Behaviour*, 1(1), 1-7. <https://doi.org/10.1038/s41562-016-0001>
- Davidson, P., & Spinoulas, A. (2015). Autonomous vehicles: what could this mean for the future of transport? *Australian Institute of Traffic Planning and Management (AITPM) National Conference*.
- Evans, K., de Moura, N., Chauvier, S., & Others. (2020). Ethical Decision Making in Autonomous Vehicles: The AV Ethics Project. *Science and Engineering Ethics*, 26(4), 3285–3312. <https://doi.org/10.1007/s11948-020-00272-8>
- Faulhaber, A., Heesen, J., & Merat, N. (2018). The Role of Behavioral Rules in Autonomous Vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58, 681-692. <https://doi.org/10.1016/j.trf.2018.06.005>
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- Gibbs, J. C. (2003). *Moral development and reality: Beyond the theories of Kohlberg and Hoffman*. Sage Publications.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029-1046. <https://doi.org/10.1037/a0015141>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55-130. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Goodall, N. J. (2014). Machine ethics and automated vehicles. In *Road Vehicle Automation* (pp. 93-102). Springer, Cham.
- Gurney, J. K. (2017). Crashing into the unknown: An examination of crash-optimization algorithms through the two lanes of ethics and law. *Albany Law Review*, 79, 183-267.



- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Harsanyi, J. C. (1977). *Rule utilitarianism and decision theory*. In *Utilitarianism and Beyond* (pp. 84-111). Cambridge University Press.
- Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PloS One*, 7(8), e42366. <https://doi.org/10.1371/journal.pone.0042366>
- Kant, I. (1785). *Groundwork of the Metaphysics of Morals*. (Reprint, 1997). Cambridge University Press.
- Lin, P. (2016). Why ethics matters for autonomous cars. In *Autonomes Fahren* (pp. 69-85). Springer Vieweg, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-662-45854-9\\_4](https://doi.org/10.1007/978-3-662-45854-9_4)
- Maxwell, B., & Narvaez, D. (2013). Moral Foundations Theory and moral development and education. *Journal of Moral Education*, 42(3), 271-290. <https://doi.org/10.1080/03057240.2013.817328>
- Shah, H., Rahwan, I., & Miller, J. (2022). Pragmatic boundaries in moral psychology: How contextual factors and emotional connections influence ethical decision-making. *Ethics and Information Technology*, 24(2), 189-205. <https://doi.org/10.1007/s10676-021-09584-9>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications.
- Thompson, J. J. (1985). The trolley problem. *Yale Law Journal*, 94(6), 1395-1415.

