Tampere University

Daniel Sell

# Ethical decision-making in autonomous vehicles:
## An interdisciplinary case for mandatory utilitarian ethics settings in autonomous vehicles

Human Computer Interaction Degree Programme in Computer Sciences
Master's Thesis
June 2023

## ABSTRACT

Daniel Sell: Ethical decision-making in autonomous vehicles: An interdisciplinary case for mandatory utilitarian ethics settings in autonomous vehicles
Master's Thesis
Tampere University
Human Computer Interaction Degree Programme in Computer Sciences
June 2023

---

This thesis will attempt to demonstrate that there is an ever increasing ethical urgency in the world, in large part due to technological evolution. We no longer have the luxury of 'being on the fence' concerning certain ethical problems. We are soon running out of time. These problems need solutions posthaste.

This thesis will attempt to demonstrate that ethical settings in autonomous vehicles (AVs) should not be chosen by the users according to their own personal preferences, rather mandatory ethics settings (MES) in AVs should be an industry-wide standard. This thesis claims that personalized ethical settings (PES) like the interactive "ethical knob" shouldn't be optional in most AVs.

Furthermore, this thesis will claim that the utilitarian ethical setting is in the most viable and beneficial MES for all in the long run, not just AV occupants, but also pedestrians, bystanders, cyclists, and other motor vehicles. This will also demonstrate that the utilitarian MES is also better for the environment, economy, sustainable development, social equality, public trust and acceptance.

This thesis will present the case for the utilitarian MES in AVs from interdisciplinary perspectives. This case will be examined from multiple different disciplinary perspectives; i.e. these issues will be inspected through different lenses. Multiple scientific, peer-reviewed studies, MATLAB computer simulations, and MIT statistics, will be referenced to provide support for the claims of this thesis.

In an attempt to address the intrinsic ethical dilemmas of this issue, multiple ethical standpoints will be contrasted and compared to each other. To base the claims of this thesis, we will also take into consideration, among others, human nature, philosophy, psychology, economics, and the environment.

The conclusion of this thesis is that while none of these cases necessarily proves that the utilitarian MES is the best ethical setting for AVs, through these interdisciplinary lenses, the utilitarian MES appears to be the best option, as it passes most of the tests with the least amount of collective and collateral damage.

Keywords: human, technology, interaction, HTI, HCI, interdisciplinary, philosophy, psychology, ethics, ethical decision-making, autonomous vehicles (AVs), mandatory, personal, ethical settings, game theory, MATLAB, simulation, moral machine, prisoner's dilemma, utilitarianism, HTI, HCI

Search words: ethics, ethical decision-making, autonomous vehicles, utilitarianism

Tampere University

## TABLE OF CONTENTS

## 1 Introduction

This thesis claims that due to technological evolution, there is an ever increasing ethical urgency in the world. We no longer have the luxury of 'being on the fence' concerning certain ethical problems. We are soon running out of time. These problems need solutions posthaste.

This thesis claims that humanity has had the luxury to indulge in moral relativism, pluralism, uncertainty, and moral ambiguity. However, the arrival of technological advances, such as machine learning, artificial intelligence (AI) and AVs, require us to address and resolve ethical issues that have heretofore remained unresolved. As a civil society, we seem to be at a moral intersection. This thesis asks the questions – Is humanity on the right path or are we headed for an avoidable ethical collision? – Could the rapidly increasing pace of technological evolution be leading humanity toward a rapidly approaching "moral deadline" or a paradigm shift?

In an attempt to address and start resolving at least one of these ethical problems, dilemmas relating to the ethical decision-making processes of AVs in unavoidable collision situations (UCSs). The thesis researches the questions:

- In UCSs, which ethical decision-making setting is the best?
- Should AVs be equipped with an "ethical knob" enabling users to choose their own personal ethics setting (PES), or should there be a standard mandatory ethics setting (MES) industry-wide?

These questions will be examined and analyzed through interdisciplinary lenses. The disciplines utilized as lenses in this examination are those of philosophy (namely ethics), psychology (namely behaviorism, game theory, and buyer's mentality). and civics (namely the role of citizens and policy makers). While this thesis will use different disciplinary lenses to approach and examine these issues, the lenses are not employed evenly. Some lenses will have a greater magnification than others. Through some lenses, the focus will be narrow and comprehensive, and through other lenses the focus will be in broader contexts to examine interrelation, correlations, associations, and even possible causations.

Through the *Ethical Lens*, this thesis will examine the ethical ideals that individuals should strive for and the moral value of actions according to those different ethical theories. The ethical lens is more theoretical and focuses on moral agency and grapples with the different applicable ethical theories to the AV ethical setting dilemma. This master's thesis, especially the Ethical Lens, is an extension of the author's previous bachelor's thesis "The ethical reasoning and programming logic behind the ethical decision-making processes of autonomous vehicles". (Sell, 2023)

To answer the research questions of this thesis, the primary emphasis will be on the *Ethical Lens*. As the subject matter is about ethical AV decision-making, obviously the ethical aspect is significantly relevant. Also, we as a civilized society, have an ethical and moral duty to do the right thing when evidence is brought to light. Part of the aim of this thesis is to inspire and promote stakeholders (and realize that we are all stakeholders) to become learned on the subject, take interest, and take a stand or at least cast a vote so that we can come to a consensus and move forward on these issues and make progress.

Through the Behavioral Lens, this thesis will examine an individual's actions in practice. The behavioral lens approaches the issues of this thesis from a more pragmatic perspective. While the ethical lens focuses on ideals, this behavioral lens focuses on an individual's actual everyday behavior, attitudes, and tendencies according to game theory theorems, buyer's mentality – namely a customer's willingness to pay (WTP). The secondary emphasis will be on the Behavioral Lens because human nature and behavior is, as peer-reviewed literature as suggested, inextricably and undeniably connected to an individual's actual intentions, attitudes, and actions within society.

Through the *Civics Lens*, this thesis examines civics, group dynamics, and combines the information obtained through the previous lenses. This lens will examine social matters such as civics and suggest theoretical but justifiable social and regulatory implications based on the information obtained through the focuses of these lenses. This lens will try to get an 'overlapping perspective'. Through this overlapping perspective, it will be easier to form ideas of an interdisciplinary consensus on these matters. Through these lenses it will be clear how each aspect relates to others and mutually affects and reinforces the claims of this thesis that the utilitarian MES is the optimal choice and that the utilitarian ethical setting should be mandatory industry-wide.

The final emphasis will be on the Civics Lens. This lens will attempt to paint a picture by which we will be able to build the most civilized society. In a way the Civics lens encapsulates and embodies the knowledge obtained in the previous lenses and demonstrates correlation, possible causation, and suggests regulatory incentives and policies based on these facts. Through the civics lens it will be clear how the utilitarian MES is best and why it needs to be mandated and regulated by third parties.

Then there will be conclusions of the findings of each lens. After the lenses, there will be a discussion about the combined findings of each of the lenses and how they relate to each other and reinforce the claims of this thesis. After the discussion there will be a conclusion that the utilitarian MES is the most ethical, optimal, universally applicable, and sustainable ethical setting when viewed through these lenses. At the end, will be a final conclusion of the thesis as a whole, personal social commentary on the heightened ethical exigency in the world and our personal role in resolving these issues.

## 2 Technological Evolution & Rising Ethical Urgency

The theory of ‘technological evolution’ is that society is radically transformed through advancing technological progress and development (Böhme & Stehr, 2012). In 1993, technology was defined as "a material entity created by the application of mental and physical effort to nature in order to achieve some value" by Radovan Richta, and according to Richta, technology evolves in three stages: tools, machine, automation (Bloomfield, 1993). This thesis is mostly focused on the third stage of technological evolution – automation.

Technology alone, is neither, good, bad, nor neutral as Melvin Kransberg postulated in the first of his, now famous, ‘six laws of technology‘ (Kranzberg, 1986). Technology can be a powerful catalyst, but it is just a tool. This thesis will demonstrate, like Kranzberg’s fourth law states, that while technology is a significant and considerable factor in many public and social issues, ‘nontechnical factors take precedence in technology-policy decisions’ (Kranzberg, 1986). As a civilization, the choices we make and the use of these tools, ethical or not, will define us.

Due to the process of *Technological Evolution,* there seems to be a noticeable increase in the burden on our moral load. There can be felt, in the world, an ‘ethical pressure’ accumulating because of advancing technological evolution. There is a heightened ethical urgency in the modern world to address and answer questions that have, as of yet, remained unanswered. We have increasing pressure to address ethical issues that have hitherto been contentiously debated or left open-ended.

For the sake of understanding the pervasive scope of these issues, here are some of the many ways in which there is a heightened urgency to answer ethical dilemmas: privacy, data rights and protection, intellectual property and ownership, impact on employment and labor, social and economic inequality, unemployment and human dignity, transparency & explainability, cybersecurity & hacking risks, informed consent in AI systems, autonomy in warfare, environmental impact, digital addiction & mental health, accountability and responsibility, algorithmic bias and discrimination, and implications for decision-making in autonomous processes, such as AVs.

The above examples highlight the diverse range of ethical challenges emerging from technological advancements. Addressing these issues requires interdisciplinary collaboration, public dialogue, and a commitment to ensuring that technology aligns with human values and societal well-being.

In an attempt to address and start resolving at least one of these problems, the ethical dilemmas relating to an AVs ethical setting in an unavoidable collision situation (UCS) will be examined and an answer provided. The thesis attempts to determine which ethical decision-making setting is the best. This thesis also attempts to determine whether AVs she be equipped with an "ethical knob" enabling users to choose their own personal ethics setting (PES), or should there be a standard mandatory ethics setting (MES) industry-wide?

## 3 Ethical Decision-Making of AVs

Before we can ethically program AVs to make ethical decisions, we need to come to an overlapping consensus on our shared ethical preferences. Thus far, humanity has been lost in ambiguity and vague notions such as moral relativism and pluralism, and uncertainty – resulting in a considerable amount of disaccord and contentiousness. However, due to the increasingly rapid technological advances in the world, such as automation, machine learning, self-driving cars, and A.I., we are encountering increasingly urgent moral dilemmas and we need to come to a consensus as certain ethical conundrums need prompt real-world solutions.

As previously mentioned, there is an urgent need to address these ethical dilemmas and to develop guidelines for the responsible use of AVs. While it is necessary to have guidelines and criteria for ethically responsible use of AVs (Goodall, 2014), developing these guidelines for ethical decision-making processes of AVs is no simple task. Many ethical dilemmas, quandaries, and conundrums have plagued and perplexed humanity for eons and have divided us by our constitutions. Thus far in our technological and civil evolution, humanity has yet to come to a collective consensus on the dilemmas related to the ethical decision-making processes and ethical settings of AVs.

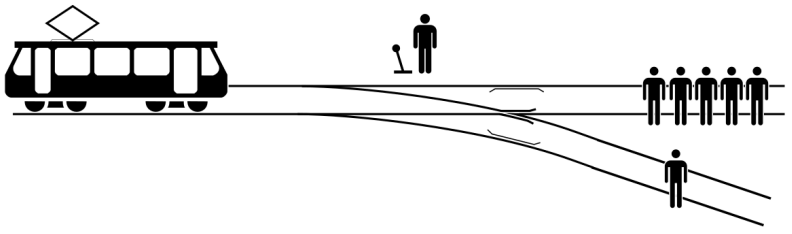


Figure 1. The trolley problem (Thomson, 1976)

With the arrival and adoption of these technically advanced AVs, the classic philosophical thought experiment, the trolley problem (Figure 1), is no longer just a hypothetical thought experiment but a real

question in need of a definitive answer (Thomson, 1976). In the event of an unavoidable collision scenario (UCS), where an AV must choose between hitting a pedestrian or swerving and potentially causing harm to the vehicle's occupants, what should the AV be ethically programmed to do? Should the AV always protect its occupants at all expense or should the AV always try to minimize harm and fatalities?

This new turn in the trolley problem from thought experiment to serious question is proof that there is a heightened ethical urgency in the world, but what is the best ethical programming for AVs in an UCS in which human lives are at stake? Should individuals have the option for a customizable personal ethics settings (PES) or should there be mandatory ethics settings (MES) in AVs? How can we ethically program automated processes like AVs with a universally applicable ethical preference, or preferences, without sharing a collective consensus on these matters? How can we answer these previously unresolved ethical dilemmas? This thesis will attempt to answer such questions.

### *3.1 The Dilemma of MES vs PES*

Some of the main disputed issues in the realm of ethical decision-making processes in AVs is the debate between the validity and viability of ethical candidates for an AV's ethical settings, and whether those ethical settings should be a personally chosen ethics setting (PES) or a mandatory ethics setting (MES). There are many valid arguments for both ethical options. Many of the disputed arguments between the PES vs MES issue lie in liability and accountability on behalf of the AV users, shareholders, manufacturers, programmers, designers, industry and policymakers alike. This thesis will address these issues and propose the most valid ethical setting based on the conclusions that can be made when viewed through the interdisciplinary lenses utilized in this thesis

In the next sections, we will introduce the PES and MES, and examine the controversy between the two options. The particular disputed arguments for and against each ethical setting will be examined later through the interdisciplinary lenses. Through the focus of these interdisciplinary lenses, we will be able to form a verdict between the two. There will be further examination of applicable ethical candidates in the Ethical Lens.

#### 3.1.1 MES - Mandatory Ethics Setting

Individuals can have differing opinions for public issues compared to personal issues (Joo & Kim, 2022). While they'd prefer everyone else have utilitarian settings in their AV, they also personally prefer an AV with selfish settings for themselves (Gogoll & Müller, 2017; Joo & Kim, 2022). As individuals, people are susceptible, perhaps even prone, to selfish behavior and weak adherence to moral principles when subjected to temptation. As stated in the bible, 'the spirit is willing but the flesh is weak'. (Matthew 26:41) To err is

human. For these reasons and many others are of the opinion that AV ethical settings should be mandatory for the greater good and mandated by third parties such as governmental regulation and industrial standards.

Some individuals are of the opinion that AVs shouldn't be ethical "tabulae rasae" (i.e. blank slates) to be used at each user's own discretion or choosing the ethical route of inaction, like bumper cars on the road (Gogoll & Müller, 2017). While the removal of the 'ethical knob' will lessen the human-computer interaction on the surface, proponents of the MES actually believe they are adhering to our core human beliefs by embedding our core human beliefs into the programming of AVs for the benevolent reason of maximizing the overall benefits to society. Others are of a different opinion, they believe that AV users should have the autonomy and right to choose their own PES, as is the case with the Ethical Knob.

### 3.1.2 PES - The 'Ethical Knob'

It has been suggested that AVs be equipped with an "Ethical Knob". The ethical knob is an interactive and adjustable knob that gives the AV user the freedom of choosing the AVs ethical setting to their own PES (Contissa et al., 2017). As seen in Figure 2, the ethical knob has three settings, an altruist setting with the preference for sparing other third parties, an impartial setting that equally considers self and others without preference, and an egoist setting with a clear first party preference for self-preservation and complete disregard for others.

Using the altruistic setting, the importance of other people's lives outweighs the importance of the AVs own occupants' lives. Therefore, the AV would always sacrifice its own occupants in order to save other third party individuals, such as passersby and pedestrians. Alternatively, using the egoist setting, the importance of the AVs own first party lives outweighs other third party lives and would always sacrifice passersby or pedestrians rather than its own occupants. With the egoist setting, first party preservation preferences can be extended to include family and kin. (Contissa et al., 2017)

Utilizing the impartial setting, there is an equal preference between the first party lives of AVs own occupants as the lives of other individuals. One could say of the impartial setting that equal preference is no preference. However, impartial does not mean indifference or inaction in an UCS. The impartial ethical setting would prioritize minimizing overall casualties regardless of first or third party affiliation, much like the tenet of utilitarianism (Contissa et al., 2017). Given equally similar circumstances, such as identical number of occupants per AV involved in an UCS, there might be a preallocated preference for first or third parties, or even a random preference between the involved parties. (Contissa et al., 2017)

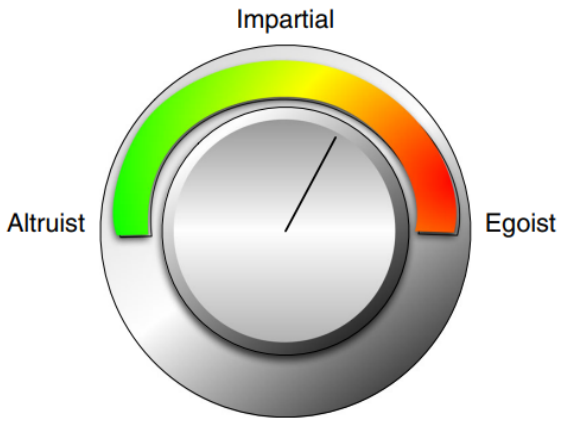



Figure 2. The ‘Ethical Knob’ (Contissa et al., 2017)

One of the many benefits of the ethical knob is that it provides adjustable ethics settings that respects the AV user's autonomy by devolving this decision-making process back to its original source, the user, along with the associated accountability and liability in both the moral and legal sense (Contissa et al., 2017). A major ethical dilemma is that of responsibility, and who should be held accountable for legal liability. To whom should the responsibility and accountability be assigned: the AV manufacturers, the programmers, the engineers, captains of industry, policy makers? In this sense, the ethical knob provides a solution to one of the more significant and far reaching ethical dilemmas in the realm of AV decision-making, by putting the burden on the user.

There could be the option to either have an AV model with a permanent preset PES individually chosen upon purchase of an AV or the option for a model with a constantly adjustable PES that would enable the AV user to switch between ethical settings, such as when there is a ‘baby on board’ or other extenuating circumstance such as AVs containing VIPs. (Contissa et al., 2017)

There are more nuanced and intricate and interconnected aspects to this issue. There are also other ethical viewpoints and AV ethical setting options than are mentioned in the Ethical Knob. They will be examined and discussed in the following chapters.

## 4 HTI is Interdisciplinary

Human-technology interaction (HTI), as a field of study, is the research in design and use of computer technology and the interaction between man and machine, otherwise called human-computer interaction

(HCI), man-machine interaction (MMI), human–machine interaction (HMI), or computer-human interaction (CHI). HTI is intrinsically an interdisciplinary field according to the scientific textbook *Research Methods for Human-Computer Interaction* (Cairns & Cox, 2008). Therefore, one must approach the subject from multiple angles and draw from multiple different disciplines and areas of expertise. In other words, one must inspect HTI issues from different disciplinary perspectives, i.e. through different *Lenses*, to achieve a comprehensive and all-encompassing understanding of the subject matter at hand.

Focused through these multidisciplinary 'lenses', this thesis will show that the utilitarian ethical setting in AVs is the optimal ethical option. This thesis will further claim, and demonstrate, that the utilitarian ethical setting should be mandatory in AVs industry-wide, pending exceptions (Gogoll & Müller, 2017). Utilizing these interdisciplinary lenses, it is possible to attain a perspective that is simultaneously broad (multiple disciplinary approaches) and narrowly focused (case for the utilitarian MES in AVs). With a wider and clearer understanding/picture, we can better infer from, extrapolate on, and form better informed verdicts and consensuses on these contemporary, complex, and controversial issues.

In the next lens, the philosophical aspects of this issue will be comprehensively examined through the Ethical Lens in Chapter 5; the psychological aspects of this issue will be examined in the Behavioral Lens in Chapter 6; the rights and duties of citizens – all citizens, from the general public to the policymakers – will be examined in the Civics Lens in Chapter 7. Through these lenses, it is possible to obtain a comprehensive perspective and make the informed conclusion that the utilitarian MES is the optimal option. This will be demonstrated in the Discussion in Chapter 8 and the Conclusion in Chapter 9.

## 5 Ethical Lens

In essence, determining the most appropriate ethical setting for an AV is reduced to an ethical one. Differing ethical theories provide alternative perspectives, each prioritizing certain fundamental principles (tenets) and concerns over others. This brings us to the philosophical '*Ethical Lens'*.

In the future, when AV technology works as well in practice as it is supposed to in theory, AVs will reportedly be orders of magnitude safer and more reliable than the average human driver today. AVs can remove the risk of human error and reduce the risk of collision by 90% compared to traditional vehicles with human drivers (Litman, 2014). However, reality is "fuzzy" and foreseeing all possible road events is unrealistic (Nyholm & Smids, 2016). Unavoidable events and situations like accidents are an inescapable

part of our reality. We need an ethical framework and guidelines for an AV's appropriate ethical behavioral reaction in these worst case scenarios.

There are many ethical frameworks from which to approach this issue. There are vastly different metaethical considerations and ethical approaches to this issue, such as ethical pluralism, moral relativism, and subjective personal ethics that are based more on belief, intuition, and emotion than rationality. For example, according to the theory of emotivism, ethical statements do not assert factual propositions but rather express emotional attitudes (Stevenson, 1937). In other words, all justifications are, fundamentally, no more than post hoc explanations based on emotions rather than rationality. Moral relativism acknowledges that ethical judgments can vary depending on cultural, individual, subjective, or contextual factors. Ethical pluralism recognizes that there may be multiple simultaneously valid ethical considerations and values.

However, this lens will focus on the traditional normative ethical theories/settings normally considered relevant and utilized in peer-reviewed scientific research and literature on the subject. (Kumfer & Burgess, 2015; D'Souza et al., 2022)

### *5.1 AV Ethical Settings*

An AVs ethical settings determine the AVs course of action in an UCS. Three examples of UCSs an AV may be faced with is shown in Figure 3. In example A, the AV must make the ethical decision between killing one pedestrian or multiple pedestrians. In example B, the AV must make the ethical decision between killing a single pedestrian or sacrificing the AV's occupant. In example C, the AV must make the ethical decision between killing multiple pedestrians or sacrificing the AV's occupant.

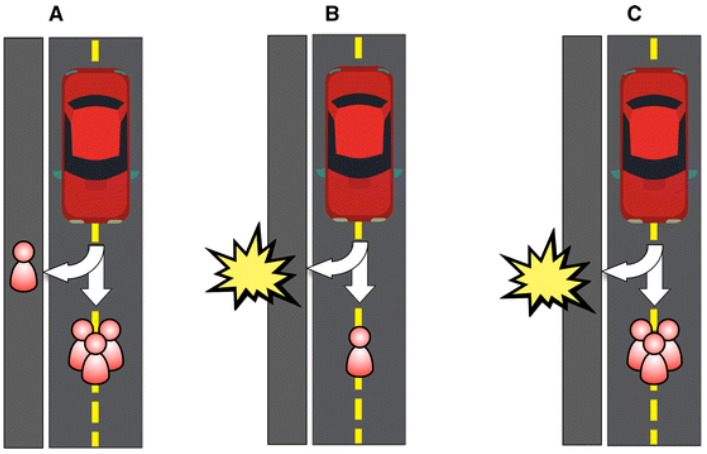



Figure 3. Three possible ethical dilemmas faced by AVs in UCSs (Contissa et al., 2017)

To make these ethical decisions an AV must rely on the ethical setting chosen by the user or installed by the manufacturer. In no particular order, this thesis will examine five of the main normative ethical theories

adopted as ethical setting candidates in AVs. The main AV ethical setting candidates in peer-reviewed literature on the subject are: deontology, virtue ethics, utilitarianism, egoism, and altruism (Kumfer & Burgess, 2015; D'Souza et al., 2022). Listed below are brief summaries of each ethical viewpoint and how each ethical setting would dictate the AVs priorities and behavior accordingly in an UCS.

### 5.1.1 Deontology

Deontological ethics emphasizes prioritizing moral character and adherence to rules, obligations, and duties regardless of the consequences (Waller, 2005). According to deontology, living a good life involves being a vigilant moral agent that develops and practices virtuous behavior consistently, and follows moral principles that would be agreed upon by all rational people, such as Kant's Categorical Imperative.

According to Kant, the main proponent of deontology, an act must be done dutifully, but it must also be the right act, and determining the 'rightness' of an act is only possible through rationality. Kant believed our wills are affected by, but not determined by bodily desires. He also believed rationality was unique to humans. He placed rationality as the deciding factor in determining an action's moral worth. According to Kant, rationality, the ability to reason, could only be attained through education and interaction with a civil society (Kant, 2003). Rationality is the main criterion in his formulation of *The Categorical Imperative*.

Kant came up with a single set of three normative principles, known as the 'categorical imperative', for which all the principles of action can be tested against to see if they are consistent with the demands of Kantian morality. The three formulations of the categorical imperative are:

1. "Act only according to that maxim whereby you can at the same time will that it should become a universal law."
2. "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end."
3. "Therefore, every rational being must so act as if he were through his maxim always a legislating member in the universal kingdom of ends" (Kant, 1785).

Simply put, the Categorical Imperative states that one should only act in the manner in which one would wish everyone to act all of the time. Kant's deontological Categorical Imperative and its relationship to societies adoption of the utilitarian MES and will be mentioned later in the conclusion.

From a deontological perspective, this would mean an AV with the deontological ethical setting would always adhere to and follow a predetermined set of rules and prioritize certain principles regardless of the

consequences For example, it might prioritize adherence to the law, virtuous behavior like doing no harm or minimizing the number of casualties.

**Criticisms of the deontologist ethical setting in an UCS**

Deontological ethics relies on predetermined rules or principles that must be followed regardless of the consequences (Waller, 2005). Critics argue that this rigidity may not always allow for nuanced decision-making in complex and unpredictable situations. In the context of AVs, strict adherence to a predefined set of rules might not adequately address the dynamic nature of traffic scenarios and the need for context-specific decision-making.

Deontological ethics often disregards the potential consequences or outcomes of an action (Waller, 2005). In the case of AVs, this means that the focus is solely on following predetermined rules or principles without considering the potential harm or benefits that could result from different actions. Critics argue that such an approach may lead to suboptimal or morally questionable decisions, as it fails to prioritize the overall well-being or minimize harm in real-world scenarios.

Deontological ethics treats all moral rules or duties as equally important and absolute (Waller, 2005). Critics argue that this fails to account for the potential conflicts between different moral principles. In the context of AVs, situations might arise where following one moral rule (e.g. protecting occupants) might directly conflict with another moral rule (e.g. minimizing harm to pedestrians). The lack of a clear hierarchy or decision-making framework can make it challenging to resolve these conflicts ethically.

Deontological ethics is often based on existing moral frameworks or established rules. However, AVs may encounter novel situations that fall outside the scope of existing rules or ethical systems. Critics argue that deontological approaches might provide limited guidance or fail to address the unique challenges and ethical dilemmas posed by autonomous driving technology.

Deontological ethics tends to be static and resistant to change (Waller, 2005). Critics argue that as societal values and norms evolve, ethical frameworks should also adapt to reflect these changes. AVs present an opportunity to redefine ethical standards and expectations. However, a strict deontological approach might hinder the ability to incorporate moral progress and adapt to evolving ethical considerations.

These criticisms do not dismiss the value of deontological ethics entirely. Rather, they highlight the challenges and limitations of relying solely on a deontological ethical setting in the context of AVs. Integrating other ethical frameworks or considering the consequences and context in decision-making processes can help address some of these criticisms and provide a more comprehensive ethical approach.

### 5.1.2 Virtue Ethics

Virtue ethics emphasizes the development of moral character traits. According to virtue ethics, to live a good life, a virtuous person should strive to be the best version of themselves and act in accordance with their moral character. An AV guided by virtue ethics would prioritize virtues such as caution, responsibility, fairness, equality, and beneficence (Aristotle et al., 2011). The AV would aim to minimize harm to all parties involved and make decisions based on virtuous behavior. Commonly accepted virtues include, among others, prudence, justice, fortitude, magnanimity, and temperance (Aristotle et al., 2011). Some of the commonly known vices include the seven cardinal sins of pride, greed, envy, wrath, lust, gluttony, and sloth. (Burkley et al., 2018)

Virtue is found in the golden middle ground between two contrasting vices (Aristotle, 2011). The virtue of courage, for example, is the golden middle between recklessness and cowardice. Between vice and virtue, doing the right thing is often the harder thing to do. One must often choose between happiness (as instant gratification: eating a cheeseburger) and being virtuous (as delayed gratification in order to achieve a more important goal: physical fitness/longevity).

In other words, to be virtuous, one ought to take the path less taken, the road less traveled. To quote Matthew from The Bible, one ought to “enter through the narrow gate. For wide is the gate and broad is the road that leads to destruction, and many enter through it. But small is the gate and narrow the road that leads to life, and only a few find it.” (King James Bible, 2017) Or to use pop culture references from the music industry, there is a “stairway to heaven” and a “highway to hell”, and to be virtuous one ought to “walk the line” between excess and deficiency.

#### Living virtuously & role virtues

Similar to the sentiment in Kant’s categorical imperative, virtue ethics asserts that the right action will be that chosen by a 'virtuous agent’ (Hursthouse & Pettigrove, 2022). Additionally, virtue ethicists claim that an agent’s actions are not truly virtuous until they have become habitual. A moral agent should act virtuously (eg. courageous, benevolent, reverent behavior) consistently, not only when it is convenient. (Smith et al., 2008)

Virtues specific to a particular profession are called role virtues. For example, engineers are encouraged to develop the virtues of technical competence, discernment, and autonomy. Engineers and role virtues are mentioned in this thesis because engineers are expected to hold the safety and wellbeing of the public above all else (Smith et al., 2008). With power comes responsibility and engineers are implored to use their practical expertise to protect the public and not treat humans as just a means to an end, but to also make

ethically just decisions for those less knowledgeable. Assuming this is true, engineers have an ethical duty to design AVs ethically and in a manner that is in everyone's best interest.

**Criticisms of the virtue ethics ethical setting in an UCS**

The virtue ethics setting, which emphasizes the development of moral character traits, also faces certain criticisms in the context of ethical decision-making for AVs. For example, virtue ethics does not provide clear, specific rules or guidelines for decision-making. It focuses more on cultivating virtuous character traits rather than providing a step-by-step process for ethical decision-making. Critics argue that this lack of explicit guidance can lead to ambiguity and subjectivity in determining the right course of action in specific scenarios.

Virtue ethics encompasses a wide range of virtues, such as honesty, empathy, courage, and responsibility. However, there can be different interpretations of these virtues and varying priorities among different individuals or cultures. Critics argue that this variability can lead to inconsistencies and conflicts when determining the virtuous course of action in UCSs. (Gogoll & Müller, 2017)

Virtue ethics can be challenging to operationalize in the programming and decision-making algorithms of AVs. While human drivers can rely on their moral intuitions and experiences to make virtuous decisions, translating these subjective qualities into specific instructions for machines can be complex. Critics argue that it is challenging to quantify or define virtuous behavior in a manner that can be implemented effectively in autonomous systems.

Virtue ethics may struggle to provide clear guidance when there are conflicting virtues or when virtues come into conflict with one another. In UCSs, it might be difficult to determine which virtue should take precedence over others. For example, the virtue of responsibility to protect vehicle occupants might conflict with the virtue of empathy for pedestrians or other road users. Critics argue that without a clear hierarchy or decision-making framework, it can be challenging to resolve such conflicts in a consistent and just manner.

Virtue ethics heavily relies on individual moral character and judgment. Critics argue that this subjectivity can introduce biases and inconsistencies in decision-making, as different individuals may have different interpretations of what virtues to prioritize or how to apply them in specific situations. This subjectivity raises concerns about fairness and objectivity in the decision-making processes of AVs.

While virtue ethics has its critics, it also offers valuable insights into the importance of character development and moral considerations. Integrating virtue ethics with other ethical frameworks and providing

clearer guidelines for decision-making can help address some of these criticisms and provide a more robust ethical approach for AVs.

### 5.1.3 Utilitarianism

Utilitarianism suggests that the action that maximizes overall utility or happiness is morally right. Utilitarianism holds that actions should be judged solely by their consequences, and that the best action is the one that maximizes happiness for the greatest number of people (Mill, 1863). Utilitarianism treats all individuals equally and impartially, irrespective of factors like age, gender, race, or socioeconomic status (Scheffler, 1988). In the context of AVs, a utilitarian approach would prioritize minimizing the amount of harm, such as reducing the number of casualties or severity of damage, and maximizing utility and the overall welfare of all individuals involved. It might make decisions to sacrifice the AV's occupants' own safety if it could save more lives or prevent more severe injuries.

**Criticisms of the utilitarian ethical setting in an UCS**

The utilitarian ethics setting, which prioritizes maximizing overall welfare or happiness, also faces criticisms in the context of ethical decision-making for AVs. For example, one of the primary criticisms of utilitarianism is the potential for sacrificing individual rights and fairness (Scheffler, 1988). Utilitarianism places emphasis on the overall outcome and maximizing aggregate happiness, which can sometimes lead to disregarding the rights and well-being of specific individuals or minority groups. Critics argue that this approach can overlook the importance of individual autonomy, dignity, and fairness in decision-making.

Utilitarianism relies on quantifying and comparing utilities or well-being, often in terms of happiness or pleasure. However, it can be difficult to accurately measure or compare the subjective experiences and values of different individuals. Critics argue that this challenge in assigning values to different outcomes undermines the objectivity and reliability of utilitarian decision-making. (Scheffler, 1988)

Utilitarianism may not adequately consider the individual preferences, values, and desires of individuals affected by a decision. Critics argue that decision-making solely based on aggregate outcomes may fail to account for the diversity of human experiences and the importance of respecting individual choices and autonomy.

In certain scenarios, utilitarianism might justify actions that are considered morally unjust or morally questionable. For example, in an UCS, a utilitarian approach might prioritize saving more lives by sacrificing the fewest lives, even if it means targeting a vulnerable group disproportionately. Critics argue that such

outcomes can be seen as ethically problematic and can undermine trust in AVs and the moral principles they follow.

Utilitarianism often focuses on the instrumental value of outcomes rather than considering the inherent value of certain actions or principles. Critics argue that this approach might neglect the significance of moral duties, intrinsic rights, or the moral worth of certain actions, reducing ethics to mere calculations of consequences.

It is important to note that these criticisms do not discount the value of utilitarian ethics entirely. They highlight the complexities and challenges associated with a strict utilitarian approach in the context of AVs. Incorporating other ethical frameworks, considering individual rights and values, and balancing utilitarian considerations with fairness and justice can help address some of these criticisms and provide a more comprehensive ethical approach.

### 5.1.4 Ethical Egoism

Egoism is the ethical theory that places self-interest as the central focus of moral behavior. There are different forms of egoism, such as psychological egoism (describing how humans actually behave) and ethical egoism (prescribing how humans ought to behave). Ethical egoism focuses on self-interest and personal benefit. According to ethical egoism, living a good ethical life involves acting in a way that maximizes one's own self-interest and well-being (Shaver, 2019). This perspective suggests that individuals should selfishly prioritize their own needs, desires, and happiness above others. However, it is important to note that ethical egoism does not endorse harming others or engaging in immoral actions. It acknowledges that cooperation and positive relationships can serve an individual's long-term self-interest. (Sanders, 1988)

From an egoistic perspective, an AV would simply prioritize the well-being and safety of its occupants above all else. It would act in a manner that minimizes harm to the AV’s own occupants and kin, even if it means disregarding the welfare of others.

**Criticisms of the ethical egoist ethical setting in an UCS**

The egoist ethical setting, which prioritizes self-interest and personal benefit, also faces several criticisms in the context of AVs. For example, the egoistic approach disregards the well-being and interests of others in favor of the self. Critics argue that this perspective can lead to a disregard for the safety and welfare of other individuals, including pedestrians, cyclists, or passengers in other vehicles. It can undermine the goal of creating a fair and just society by prioritizing one's own interests at the expense of others.

AVs programmed with an egoistic ethical setting might engage in risky or aggressive behavior to prioritize the safety and well-being of the vehicle occupants. This behavior could result in increased accidents or harm to other road users. Critics argue that such actions perpetuate a negative and selfish cycle, where each AV acts solely in its own interest without regard for the greater good.

An egoistic approach to AVs may have broader societal implications. It could lead to increased congestion, as each vehicle prioritizes its own efficiency and convenience over cooperative traffic flow. Additionally, an egoistic setting might discourage the adoption of shared mobility or public transportation solutions, as individuals prioritize their personal vehicles and interests above broader societal or environmental benefits.

The egoistic ethical setting raises legal and regulatory challenges. If AVs prioritize the safety and well-being of their occupants at all costs, there may be legal and ethical questions regarding liability and responsibility in the event of accidents. Establishing guidelines and regulations that balance individual interests with societal welfare becomes more complex in this context.

Implementing an egoistic ethical setting might lead to a lack of social acceptance and trust in AVs. If people perceive that AVs prioritize individual self-interest, it could undermine public confidence in the technology. Trust and cooperation among road users are essential for the safe and widespread adoption of AVs.

It is worth noting that the criticisms listed above do not negate the value of self-interest altogether. The concerns highlight the potential risks and ethical implications associated with an extreme egoistic approach. Striking a balance between self-interest and consideration for others is crucial in developing responsible and ethical decision-making processes for AVs.

### 5.1.5 Altruism

Altruism is the ethical theory that prioritizes the importance of acting for the benefit of others (Zahavi, 1995). Altruism emphasizes selfless concern for the well-being of others and contrasts the selfishness of egoism. Altruism holds that living a good ethical life involves selflessly caring for and promoting the well-being of others, even at the detriment to one’s self and own interests.

The tenets of altruism contrast those of ethical egoism. According to the tenets of virtue ethics, virtues are found in the middle ground between two extremes (Aristotle, 2011). Thus, it can be inferred that neither complete egoism or total altruism is best. Rather, the virtuous path is found in the golden middle, in the fluctuating ebb and flow of selfishness and selflessness. This idea of balance through harmonious change

between interrelated opposites is similar to the concept of yin-yang in the Daoist philosophy. (Lee et al., 2009)

An AV guided by altruistic principles would prioritize minimizing harm to all parties, placing the welfare of both the occupants and other possible bystanders. It would make decisions that aim to protect and save as many lives as possible, potentially sacrificing the safety of its own occupants when necessary. (MacAskill, 2017)

**Criticisms of the altruist ethical setting in an UCS**

While the altruistic ethical setting, which prioritizes the well-being of others above all else, may seem commendable, it is not without its criticisms. For example, one significant criticism of the altruistic ethical setting is that it may require sacrificing the safety of the occupants of the AV in order to prioritize the well-being of others (MacAskill, 2017). This raises concerns about the fairness and protection of individuals who have entrusted their safety to the vehicle. Critics argue that it is unfair to put the lives of occupants at risk for the sake of strangers, as occupants have a reasonable expectation of being protected by the vehicle they are traveling in.

Critics of the altruistic approach argue that it may overlook the importance of moral agency and personal responsibility (MacAskill, 2017). If the vehicle always prioritizes the welfare of others, it may relieve individuals of their responsibility to drive safely and make ethical decisions themselves. This can lead to a moral hazard where individuals become more careless or reliant on autonomous systems, potentially contributing to a decline in personal accountability.

The altruistic ethical setting raises questions about who should be considered "others" and how their well-being should be prioritized. The concept of "others" can be vague and can include a wide range of individuals, such as pedestrians, passengers in other vehicles, or even animals. Determining the hierarchy of priorities and the extent to which different lives should be valued becomes a complex and potentially contentious task.

Critics argue that an altruistic ethical setting can be exploited by individuals who intentionally put themselves in harm's way to manipulate the behavior of AVs. For example, a person might step in front of a vehicle to force it to prioritize their safety, potentially leading to ethical dilemmas and unsafe conditions on the road.

The altruistic ethical setting assumes a universal prioritization of the well-being of others. However, cultural and ethical values can vary significantly across different societies and individuals. What is considered

morally right or wrong can differ based on cultural norms and beliefs. Implementing an altruistic approach without accounting for this diversity may lead to conflicts and controversies.

It's important to note that these criticisms do not invalidate the altruistic approach entirely. They highlight the complexities and challenges associated with implementing a pure altruistic ethical setting in the context of AVs. Achieving a balanced approach that considers various ethical perspectives and societal values is crucial to ensuring responsible and ethical decision-making by AVs.

### *5.2 Ethical Lens: Studies*

Now that we have examined the subject comprehensively through the ethical lens, we should have a clear understanding of the ethical issues surrounding an AVs ethical settings and decision-making processes. Next, we will examine the state of research and scientific peer-reviewed literature on this specific subject up to this point.

#### 5.2.1 The Moral Machine

The Moral Machine (Awad et al., 2018) was an experimental online platform developed by MIT to gather public opinions on ethical decisions made by AVs in hypothetical scenarios. It presented users with various moral dilemmas involving potential harm to different groups of people, forcing them to make choices that prioritize certain individuals over others. The platform also collected data on users' demographic information, such as their country of origin, to analyze how cultural backgrounds might influence AV decision-making.

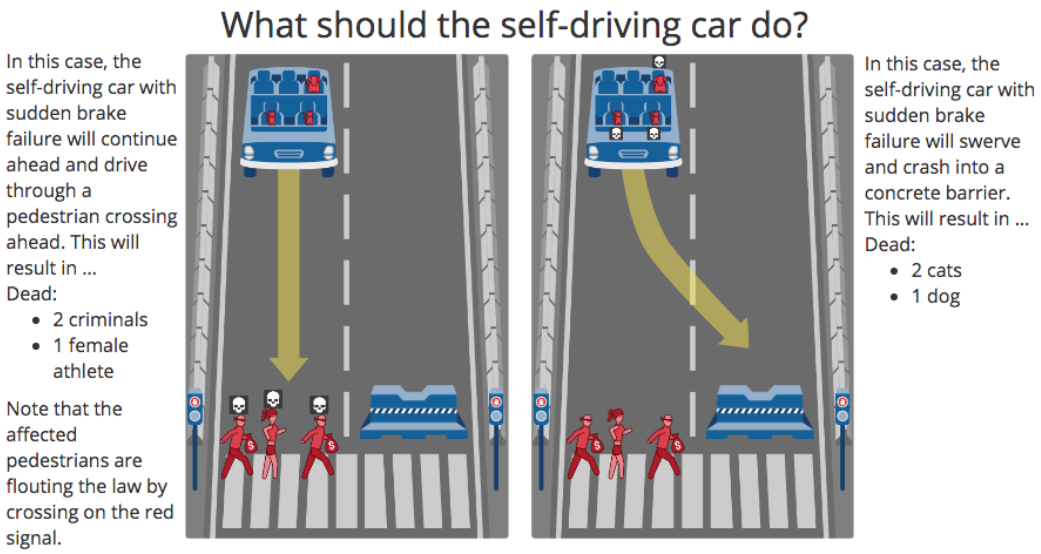



Figure 4. A Moral Machine scenario (Awad et al., 2018)

The Moral Machine project, developed by researchers at MIT, is an online platform designed to explore moral dilemmas associated with AVs. The project aims to collect data and opinions from a wide range of participants in order to better understand how people make ethical decisions in situations involving self-driving cars. (Awad et al., 2018)

To access the platform, users are required to visit the official Moral Machine website, where they can access the interactive interface and begin the decision-making process. Each participant is presented with a series of hypothetical scenarios involving a self-driving car confronted with an UCS. The user is provided with a detailed description of the situation, including factors such as the number and demographics of the individuals involved, their role in the accident, and potential consequences. (Awad et al., 2018)

Participants are tasked with making difficult moral choices and base decisions by choosing who should be sacrificed or protected in the given scenario, as can be seen in Figure 4. The choices typically involve selecting which group of individuals the autonomous car should prioritize in order to minimize harm, such as passengers inside the car, pedestrians, or specific age or gender groups. The Moral Machine project introduces variations in scenarios to examine the influence of different factors on moral judgments. These variations may include changes in the number and types of people involved, the physical characteristics of individuals, or situational factors like jaywalking or legal violations. (Awad et al., 2018)

Each test comes with 13 generated UCS from a database of 26 million different possibilities. Participants could participate and take the test as many times as they wished. After making their choices for a specific scenario, users submit their responses to the Moral Machine project. The collected data is anonymized and used for analysis and research purposes, aiming to shed light on how moral preferences differ across cultures, age groups, and other demographics. (Awad et al., 2018)

They computed the Average Marginal Component Effect (AMCE) of each attribute that the Moral Machine tested (Awad et al., 2018). The AMCE measures the average change in the outcome variable when moving from one category to another, while holding all other variables constant. These attributes tested nine factors: the value of humans lives versus those of animals, inaction versus swerving, value of the lives of the passengers versus pedestrians, sparing more lives versus fewer lives, values of gender and age, values of legal actions versus illegal ones (jaywalking), values for levels of status and even physical fitness. (Awad et al., 2018)

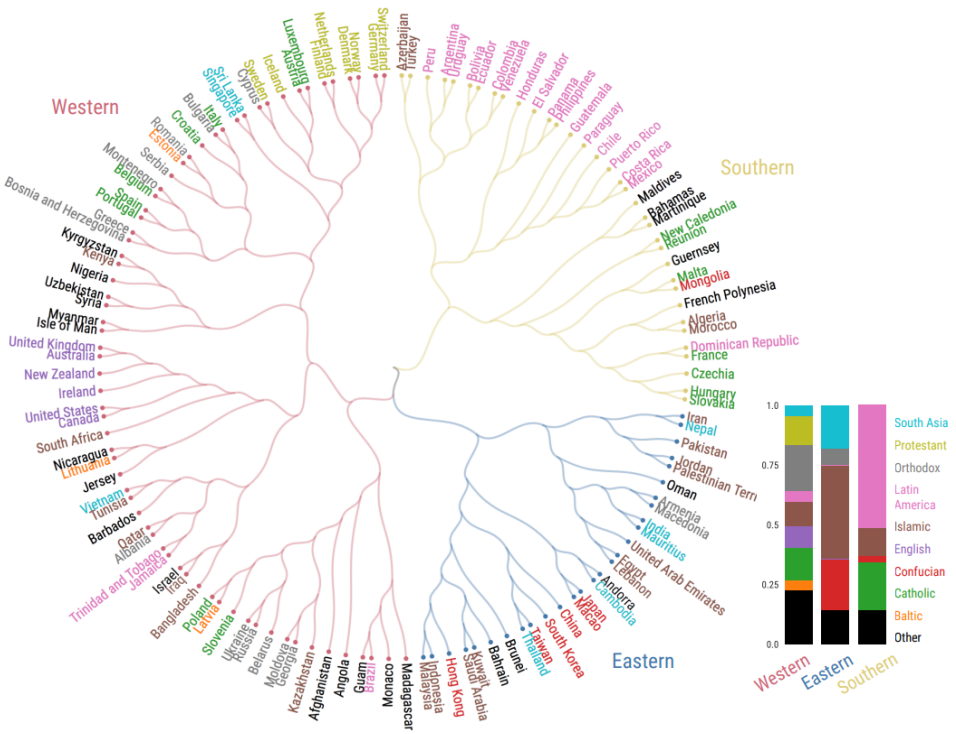
Like the trolley problem, these scenarios offer different variations of ethical ultimatums. The choices participants made helped provide a global consensus with a great degree of diversity. The information collected through the Moral Machine project was very useful in collecting a consensus and aided in further

research on the ethical decision-making dilemmas AVs, and the dilemmas autonomous processes will need to solve in the future. (Awad et al., 2018)

It is important to note that the Moral Machine project does not provide definitive answers to moral questions but instead aims to gather data to inform discussions and debates surrounding the ethical considerations of AVs. The project intends to contribute to the development of a framework that aligns public opinions with the design and decision-making processes of self-driving cars.

## Project Results

The Moral Machine was active from January 2016 to July 2020, collecting 39.61 million decisions from 233 countries. These countries, as can be seen in Graph 1, are listed and their 'cultural clusters' and affiliations are drawn out. These countries are grouped into three distinct culture clusters, Western, Easter, and Southern, as can be seen in Figure 5 and Graph 1. The ratio of the main demographic groups per culture cluster is also shown in Graph 1.




Graph 1. Country clusters, affiliations, and demographics. (Awad et al., 2018)

Upon inspection of the analysis of the data in Figure 5, the general preferences within these culture clusters appear to differ in distinctive ways. However, it's important to note that these descriptions are generalizations and do not apply uniformly to every individual within each culture cluster. Cultural attitudes and preferences are complex and can vary even within a specific culture or region. Additionally, the Moral Machine

experiment captures aggregated trends but cannot fully represent the nuanced views of every individual within a culture cluster.

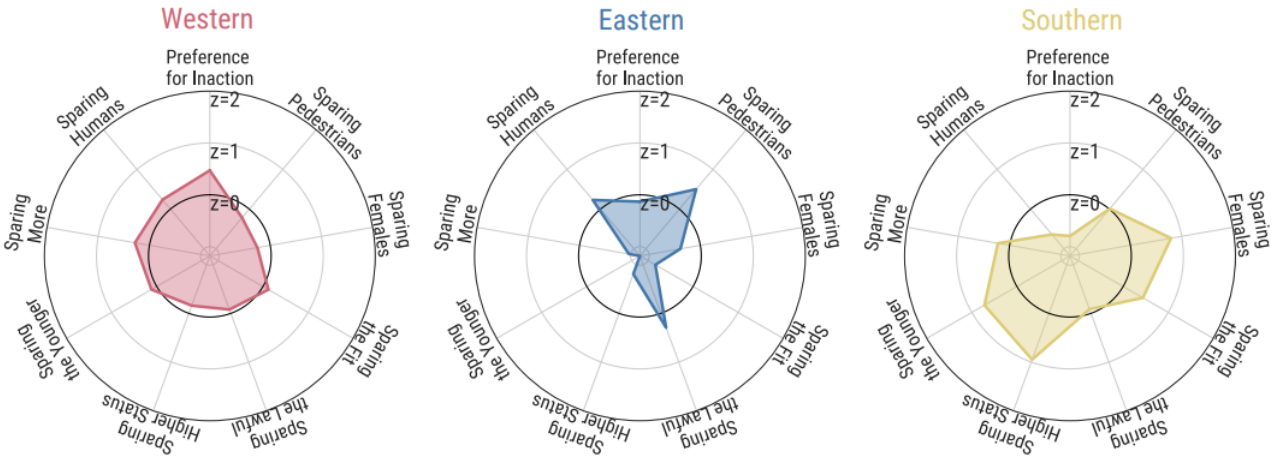



Figure 5. Cultural clusters reveal differences between the nine attributes (Awad et al., 2018)

According to the data in Figure 5, the Western cluster appears to prioritize individualistic tendencies over collective outcomes and a higher emphasis on personal autonomy and freedom of choice. The Western cluster generally shows a higher inclination to spare the lives of humans over animals and more importance placed on saving younger lives over older ones.

According to the data in Figure 5, the Eastern cluster appears to prioritize well-being and societal harmony through collectivist tendencies with a higher emphasis on social hierarchy and respect for authority and the law, higher social status, and older individuals. The Eastern cluster appears to exhibit a stronger inclination to spare the lives of humans over animals, and there appears to be a higher emphasis on saving the lives of pedestrians over passengers.

According to the data in Figure 5, the Southern cluster represents a more diverse range of countries and cultures, including Latin American and African nations, with a very strong variation in regional preferences. Preferences within the Southern cluster can vary significantly based on specific cultural, societal, and economic factors present in different regions. However, there is a higher emphasis on saving humans over animals, females over males, young over the old and those who were ‘fit’ over ‘unfit’. African cultures prioritized protecting pedestrians or vulnerable groups.

While the study provides insights into societal perspectives, it's important to note that the popularity of ethical settings can vary based on different factors such as cultural, geographical, and demographic

influences. The study aimed to capture the diversity of opinions rather than determining a single "popular" ethical setting. However, the study did reveal some global trends and patterns in public preferences.

Global statistics from the Moral Machine study point to a general preference for prioritizing humans over animals and sparing a greater number of lives and minimizing the overall number of casualties. This aligns with the principles of utilitarian ethics, which prioritize the greatest good for the greatest number of people (Awad et al., 2018). In many scenarios presented in the study, participants tended to choose options that saved more lives, even if it meant sacrificing the lives of a smaller group or occupants of the AV itself. They wanted to spare younger lives over older ones, babies being spared most often. All culture clusters shared the preference to spare pedestrians over passengers and the lawful over the unlawful. (Awad et al., 2018)

It's important to remember that these descriptions provide a general overview and do not capture the entirety of preferences within each culture cluster. The Moral Machine data provides insights into broad trends but cannot represent the views of every individual accurately. Cultural attitudes towards ethics and decision-making are complex and can be influenced by a variety of factors beyond the scope of the Moral Machine experiment. (Awad et al., 2018)

### Conclusion of Moral Machine

The Moral Machine project conducted by MIT is an important study that collected data on public preferences from millions of participants worldwide regarding ethical decisions for AVs. Analyzing these statistics helps researchers and policymakers gain insights into societal values and preferences when it comes to ethical decision-making in AVs. (Awad et al., 2018)

The Moral Machine statistics revealed that a significant portion of participants favored choices that align with utilitarian ethics, such as prioritizing the greater good and minimizing the number of casualties. This preference might indicate that a substantial segment of the public believes that AVs should make decisions based on maximizing overall welfare rather than prioritizing individual lives. While the study provides valuable insights into societal and cultural perspectives, it does not definitively prove that the utilitarian MES is the best for AVs, just that it is, on average, the global general preference. (Awad et al., 2018)

The data collected from the Moral Machine project can help inform the development of ethical frameworks and guidelines for AVs. The data provides valuable information for policymakers, researchers, and developers to consider when formulating ethical principles and decision-making algorithms. The statistics can contribute to shaping discussions around societal values, the trade-offs involved, and the potential consequences of different ethical approaches. (Awad et al., 2018)

The ethical decision-making in AVs involves complex considerations beyond popular opinion and the Moral Machine statistics. The utilitarian MES is just one ethical approach among many, and the choice of an ethical framework requires careful analysis, incorporating multiple perspectives, stakeholder engagement, societal norms, in-depth interdisciplinary discussions, and ongoing research are crucial to coming to a consensus and shaping the ethical development of AVs.

### 5.2.2 MATLAB Computer Simulated Ethical Thought Experiment

Another recent study (Kumfer & Burgess, 2015) has provided interesting data on the subject, by using MATLAB computer simulations to run a virtual ethical thought experiment similar to the trolley problem. After 100,000 simulations on the virtual outcomes of using each ethical setting in an UCS, Their study provided data of the effect an AVs ethical setting might have in real-world situations. Assuming certain parameters, based on literature (Chang, 2008), such as age and sex in relation to the probability of surviving an UCS, Kumfer and Burgess were able to calculate the average fatalities per ethics system per simulation, as can be seen in Table 1.

TABLE 1. Average Fatalities per Ethics System (Kumfer & Burgess, 2015)

| Simulation Variable | Number, by Simulation Number | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Average |
| Fatalities | | | | | | | | | | | |
| Ethical egoism | 1,029 | 995 | 1,011 | 1,033 | 1,031 | 1,033 | 978 | 988 | 1,025 | 999 | 1,012.2 |
| Utilitarian | 181 | 173 | 165 | 186 | 183 | 189 | 207 | 190 | 163 | 200 | 183.7 |
| RFP | 524 | 551 | 502 | 555 | 494 | 552 | 531 | 542 | 514 | 523 | 528.8 |
| Virtue ethics | 1,328 | 1,292 | 1,386 | 1,317 | 1,323 | 1,353 | 1,417 | 1,379 | 1,381 | 1,343 | 1,351.9 |
| Total | 3,062 | 3,011 | 3,064 | 3,091 | 3,031 | 3,127 | 3,133 | 3,099 | 3,083 | 3,065 | 3,076.6 |

Kumfer and Burgess performed ten rounds of 10,000 MATLAB simulations to achieve an average number of fatalities per ethics system. For the purposes of this thesis, only the average number of fatalities per ethics system will be considered, as seen inside the red box in the rightmost column in Table 1.

According to the data in Table 1, the utilitarian ethical setting has 2.88-7.36 times less fatalities than other ethical setting options, as seen highlighted in the green box in Table 1. In other words, There is a 288% increase in the number of fatalities when moving to the next best option from utilitarianism, RFP ethics, which is deontological and altruistic in nature for the most part. There is a 736% increase in the number of fatalities when moving to the ethical setting with the most fatalities, virtue ethics. Of these MATLAB simulations, the utilitarian ethical setting accounted for an average of only 5.97% of the total casualties of all ethics settings/systems combined. (Kumfer & Burgess, 2015)

The egoist ethical setting, preferred by individuals, potential customers, and manufacturers alike, resulted in the second largest number of casualties. The fact that there are more vehicles involved in UCSs with AVs with the egoist setting may counteract any attempt to avoid related accountability or legal liability issues. (Kumfer & Burgess, 2015) Interestingly, by risking the AV's first party single occupant's life, the casualties of RFP ethics are reduced below those resulting from the more self-centered ethical settings like the egoist and virtue ethics ethics settings. (Kumfer & Burgess, 2015)

Despite the fact that AV users may object to the idea of possibly being sacrificed for the greater good, this study's results could be grounds to base the claim that the utilitarian ethical setting is the best ethical setting and should be mandatory, when considering factors such as minimizing the number of casualties and maximizing efficiency and utility for the greatest number of individuals.

While the results obtained in this study were done with certain parameters and metrics that don't necessarily represent the whole picture, this is a very good start for further research. More simulations are needed, with broader sets of parameters and a more diverse range of AV collision scenarios**.** This study clearly demonstrates that using different AV ethical settings will have vastly different numbers of casualties in potential UCSs (Kumfer & Burgess, 2015). When implemented on a global scale, this could mean a difference of millions of lives saved or lost.

### *5.3 Ethical Lens Conclusion*

The data in the studies above suggest that the utilitarian ethical setting is the most optimal and universally applicable ethical setting. The global general preference for the utilitarian ethical setting was demonstrated in the Moral Machine project by MIT (Awad et al., 2018). The utilitarian ethical setting is also the safest, fairest, and least fatal (potentially) ethical setting as demonstrated in the MATLAB computer simulations. (Kumfer & Burgess, 2015)

While varying ethical viewpoints differ in many respects, all five ethical viewpoints examined in this thesis suggest that the utilitarian ethical setting is the best for all when seen through these lenses. Even the deontological viewpoint, often contrasted with utilitarianism, implies that the utilitarian ethical setting can be viewed as the best option. This implication can be inferred from Kant's Categorical Imperative that one should do what one wishes everyone else would do. Virtue ethics would approve of virtuous behavior such as benevolence and doing what is best for society regardless of an individual's less virtuous tendencies (Rosebury, 2021). From the viewpoint of ethical egoism, the utilitarian ethical setting serves the egoist by reducing the overall risk of collision and fatality. The apparent selfless act to risk a collision and "take one

for the team" ultimately maximizes their best interest and practically nullifies the probability of risk of collision “taking one for the team”. (Kumfer & Burgess, 2015; Gogoll & Müller, 2017)

From philosophical and ethical perspectives, the utilitarian ethical setting in AVs can be perceived as a potentially viable ethical setting for universal application. For example, utilitarianism is a consequentialist ethical theory that aims to maximize overall welfare or utility. In the context of AVs, this ethical setting seeks to minimize harm and maximize well-being for all individuals affected by AV decisions. Prioritizing overall welfare aligns with the principle of promoting the greatest good for the greatest number, which is a fundamental ethical objective in many moral frameworks.

Utilitarianism promotes impartiality and fairness by considering the interests and well-being of all individuals affected by AV decisions. It treats each individual's welfare equally, regardless of factors like social status, wealth, or personal characteristics. This approach aligns with the concept of justice and the idea that no individual's interests should be given undue preference or importance.

Utilitarian ethics provide a consistent and universalizable decision-making framework. AVs programmed with utilitarian principles can apply the same ethical standards across different situations and contexts. This consistency allows for predictability and fairness in AV behavior, promoting a sense of equality and justice among users and stakeholders.

Utilitarian ethics can help mitigate biases and subjective judgments that humans may exhibit in critical situations. AVs programmed with utilitarian ethics rely on objective factors such as minimizing harm and overall welfare, rather than being influenced by emotional responses or personal biases. This can contribute to fair and impartial decision-making on the road.

Utilitarian ethics in AVs can address practical challenges related to decision-making. AVs often encounter complex scenarios where trade-offs are inevitable, such as choosing between protecting the vehicle occupant or pedestrians. Utilitarianism provides a clear decision-making framework that considers the overall consequences, making it easier to program AVs to navigate these challenging situations.

Utilitarian ethics in AVs can be seen as a step toward ethical progress and optimization. By implementing an ethical setting that prioritizes overall welfare, AVs have the potential to reduce accidents caused by human error, save lives, and enhance transportation efficiency. This aligns with the goal of continuously improving ethical decision-making and creating a safer and more efficient transportation system.

Due to its focus on maximizing overall welfare, impartiality, consistency, and practicality, the utilitarian ethical setting should be seriously considered, if not mandatory, in AVs. By prioritizing the greater good and minimizing harm, AVs can contribute to a safer and more ethically sound transportation system.

### 5.3.1 Criticisms

However, the utilitarian ethical setting is not without its criticisms and challenges. Concerns regarding individual rights, fairness, and the inherent complexities of moral decision-making in specific scenarios must be carefully considered.

There are many possible angles from which to criticize or attempt to negate what is claimed in this thesis. Ethics is the study of human ideals, their “best intentions”, however, ‘the road to hell is paved with good intentions. “Let me help you so you don’t drown”, said the monkey, safely putting the fish up the tree. As Alan Watts stated, it could be viewed that it is conceited of us to assume we know any better. (Columbus & Rice, 2017) For example, how could we know that adding more roads would paradoxically slow down overall traffic flow (Pigou, 2002).

One of the many concerns about the utilitarian viewpoint/setting is the concern about assigning different values to different human lives. Are the lives of children and elders of equal value to the utilitarian? Saving the life of a three year old child could possibly be saving more years of life than saving the life of a sixty year old. On the other hand, perhaps the sixty year old is currently much more useful for society than the three year old. How can one base these decisions in lines of code? Are women and men of equal value? Are successful people of equal value to criminals? According to the data from the Moral Machine, there are differences in value according to the general consensus of a specific demographic (Awad et al., 2018). There is clearly room for individual interpretation and the potential for biases in the utilitarian decision-making algorithms of AVs.

Another aspect of utilitarianism that could be difficult for people to embrace can be understood through the thought experiment of the ‘utility monster’. In a utilitarian world, there could exist a cookie monster that gets immeasurably more pleasure from cookies than anyone else. This could result in one having to sacrifice one’s own pleasure at the expense of promoting a greater net gain in pleasure (Scheffler, 1988). When put in these words, utilitarianism could demand selflessness beyond the capabilities of most humans. (Smith et al., 2008)

The behavior of AVs in UCSs depends on the programming and design decisions made by the engineers and policymakers. Interdisciplinary ethical examinations provide different perspectives, and the actual

implementation may involve a combination of these approaches or other considerations based on societal norms and legal requirements.

In summary, the utilitarian ethical setting in AVs can be perceived as a viable option through the philosophical *ethical lens* due to its focus on overall welfare, impartiality, consistency, and practicality. Nonetheless, it is essential to engage in ongoing discussions and deliberations to ensure a comprehensive and ethically sound approach to the universal application of AV technology.

## 6 Behavioral Lens

Through the last lens in Chapter 5, we examined the more theoretical ethical aspect of an AV's ethical decision-making processes, and how people ideally ought to act according to the relevant normative ethical theories. In Chapter 6, through this lens, we'll examine what people do in practice, and how their everyday actions and behavior can be slightly different to, if not in complete contradiction with, their ethical ideals. Individuals can be moral agents as well as rational agents (Ben-Zeev, 1982; Geiger, 2011). However, they can also be fallible agents. (Pope, 1963)

This chapter will examine aspects of an individual's dualistic tendencies. While human reasoning is essentially logical and scientific, humans are often motivated and influenced by emotionally charged passions (Stevenson, 1937; Hart, 1996). Individuals can have mixed, and even contradicting preferences, depending on factors such as mood and context (Miller, 1993). For example, individuals can have varying stances on the same issue, depending on whether the issue is of a public or of a private matter or whether they are approaching the ethical issue from the perspective of a member of society or that of an individual. As a member of society, individuals are prone to act in a manner that benefits society. However, as an individual, people are generally prone to maximize personal benefits, especially as a customer (Joo & Kim, 2022). This brings us to game theory.

### *6.1 Game Theory*

As opposed to an ideal moral agent as discussed in Chapter 5, this chapter examines the individual as a rational and fallible agent who is prone to maximize personal gain regardless of their good intentions. Most individuals are not inclined to sacrifice themselves for individuals they don't know and who probably

wouldn’t do the same for them. Individuals want to be moral, but they “don’t want to be a sucker”. (Gogoll & Müller, 2017)

Game theory, with the prisoner's dilemma and “game of chicken” as prominent examples, can be utilized to help analyze the dynamics of strategic decision-making and provides insights into cooperative and competitive behaviors in various contexts.

If AVs are equipped with an adjustable PES as in the case with the ethical knob, most individuals, given freedom to choose one's own ethical settings, will ultimately pursue personal advantage and to maximize personal benefit. In this case, that means individuals, despite their ideals and social intentions, will have a probabilistic tendency to opt for the egoist setting in their own AV. They claim that, according to the principles of game theory, this will eventually result in a prisoner’s dilemma and a ‘crowding out of morality’. (Gogoll & Müller, 2017)

### 6.1.1 Prisoner’s Dilemma

The prisoner's dilemma is a useful concept in understanding various real-life situations where individuals or entities face similar trade-offs between cooperation and defection. It can be applied to diverse fields such as economics, politics, and social interactions. The prisoner's dilemma is a classic example in game theory that demonstrates how individual rationality can lead to suboptimal outcomes for all parties involved. (Poundstone, 1993)

In the classic prisoner's dilemma, two individuals are arrested and detained separately, with limited evidence against them. Either prisoner can either cooperate with the prisoner by not snitching or betray the other by snitching. The possible outcomes and associated penalties are as follows:

- If both prisoners stay quiet (cooperate), they both only get a moderate penalty.
- If both prisoners snitch (betray/defect), they both get severe penalties.
- If one prisoner snitches while the other stays quiet, the one who snitched will get a reduced penalty while the other gets a more severe penalty.

The dilemma arises because each prisoner faces the temptation to betray the other to avoid the severe penalty. However, if both prisoners act in their individual self-interest and confess, they end up with a worse outcome compared to if they had both cooperated. (Gogoll & Müller, 2017)

For the purposes of this thesis, the prisoner’s dilemma will be examined from the perspective of two AVs, Driver 1 and Driver 2, involved in an UCS. In the context of an UCS, the prisoner’s dilemma, when

considered as similar to the dilemma of choosing an AVs ethical setting, can be compared to choosing the utilitarian ethical setting (cooperating) versus choosing the self-interested egoist setting (defecting).

Table 2. Prisoner's dilemma (Acevedo & Krueger, 2005)

| Prisoner's Dilemma | | Driver 1 | |
|---|---|---|---|
| | | *cooperate* | *defect* |
| Driver 2 | *cooperate* | 8 / 8 | 0 / 10 |
| | *defect* | 10 / 0 | 4 / 4 |

In an UCS between two AVs that are either cooperating with (read: using utilitarian PES) and/or defecting against (read: using egoist PES), there is a chance for each driver to maximize personal gain by defecting, as can be seen in the upper right and lower left quadrants of Table 2. However, this personal gain comes at the other driver's expense (and society's), and will lead to the other driver and society suffering large losses in the long run. If one driver cooperates and the other defects, this renders a maximum possibility of 10 utility points in total (an uneven situation of 10/0 or 0/10).

There is also the chance that both AVs defect (both are using the egoist PES). This leads to a suboptimal result that leads to the least amount of potential utility points for all involved. If both drivers defect, this renders a maximum possibility of 8 utility points in total (a mutual situation of 4/4).

However, if both parties cooperate (both AVs are using the utilitarian PES), this leads to the overall most optimum result of a total of 16 utility points. (a mutual situation of 8/8). In this scenario, while each party didn't choose the option that could have resulted in the most personal gain, they both chose the option that most benefited themself without burdening the other, thus producing the most overall benefit and utility. (Acevedo & Krueger, 2005)

The current general consensus is that individuals, given the choice, would rather not use an AV that might sacrifice them in an UCS. Gogoll and Müller claim that this would, in actuality, be in their best interest for them and society as a whole. Contrary to what one would intuitively expect, using the utilitarian ethical setting and 'putting oneself at risk' actually reduces the probability of being involved in a collision (Gogoll & Müller, 2017). This same phenomenon was noticed in the results of the MATLAB computer simulations when using the RFP ethics setting. (Kumfer & Burgess, 2015)

Gogoll & Müller consider the resulting prisoner's dilemma to be sufficient basis for their claim that the ethical settings in AVs ought to be mandatory for society's own benefit. The personal incentives to defect will cause individuals to have a tendency to choose a selfish PES and this will lead to a 'crowding out of morality'. As a result, Gogall & Müller claim everybody would be better off if the ethical setting is mandatory and that the MES be enforced by a third party, such as governmental or industrial standards and regulations. (Gogoll & Müller, 2017)

Game theorists use the prisoner's dilemma as a starting point to study strategies that can lead to more optimal outcomes. One such strategy is known as tit-for-tat, where individuals eventually start to mirror the other's previous move. This strategy promotes cooperation by reciprocating positive actions and punishing defection/selfishness. In the terms of this thesis, utilizing the tit-for-tat strategy could help promote the adoption of the utilitarian ethical setting by punishing self-interested ethical settings, such as higher insurance prices or taxes for selfish AVs. (Contissa et al., 2017; Schwarting et al., 2019)

### 6.1.2 Game of Chicken

According to speculation, the initial stages of the deployment of new technologies such as AVs may be turbulent (Acevedo & Krueger, 2005). The lack of the general public knowledge of AVs may impede the public acceptance and social approval. (Joo & Kim, 2022)

If other drivers know an AV is using a particular ethical setting that will swerve to avoid collision, they may exploit that feature. Individuals may engage in ignorantly playful, reckless, and even aggressive behavior with AVs such as the "game of chicken" (Acevedo & Krueger, 2005). In worst case scenarios, this kind of reprehensible behavior could lead to disastrous consequences for individuals and society alike.

The "game of chicken" involves a situation where two parties engage in risky behavior, such as driving towards each other head-on, and the outcome depends on which party swerves first to avoid a collision. In other words, two drivers can either choose to cooperate or defect as in the prisoner's dilemma. Both drivers can maximize personal and society gain by cooperating (8/8 = 16 total maximum utility points, even and fair). However, each driver has an incentive to attain maximum personal gain by defecting, assuming the other driver cooperates (10/4 or 4/10 = 14 total maximum utility points, uneven and unfair).

Table 3. "Game of chicken" (Acevedo & Krueger, 2005)

Game of Chicken

| | | Driver 1 | |
|---|---|---|---|
| | | *cooperate* | *defect* |
| Driver 2 | *cooperate* | 8 / 8 | 4 / 10 |
| | *defect* | 10 / 4 | 0 / 0 |

According to the "game of chicken" theorem, drivers will have a probabilistic tendency to defect in order to avoid exploitation. Unfortunately, this leads to suboptimal outcomes that are individually and socially undesirable, especially if both defect (0/0 = 0 total maximum utility points, fair and even but unwise), as can be seen in the lower right quadrant of Table 3. (Acevedo & Krueger, 2005)

### 6.1.3 Nash Equilibrium

The Nash equilibrium is another decision-making theorem from within the realm of game theory that can be applicable to the dilemma of whether or not the utilitarian ethical setting should be the global standard MES for AVs. The Nash equilibrium states a player can achieve the desired outcome by not deviating from their initial strategy. In the Nash equilibrium, each player's strategy is optimal when considering the decisions of other players. A Nash equilibrium occurs when each participant in a game, knowing the strategies chosen by others, has no incentive to unilaterally deviate from their chosen strategy. In the context of AVs, the Nash equilibrium can also be applied to the ethical decision-making process of individual AVs in UCSs. (Schwarting et al., 2019)

Nash Equilibrium is a game theory concept that determines the optimal solution in a non-cooperative game in which each player lacks any incentive to change his/her initial strategy**.** A game can have a Nash equilibrium even though neither player has a dominant or dominated strategy. In fact, every finite game has a Nash equilibrium, possibly in mixed strategies (Aumann, 1985). The "game of chicken" above is an example of a game with no dominant or dominated strategies but which has a Nash equilibrium.

There are many ways in which the Nash equilibrium theorem can be applied when considering what could probabilistically be the overall most optimal MES for society. For example, in a Nash equilibrium, each participant chooses their strategy in a way that maximizes their own utility given the strategies chosen by others (Schwarting et al., 2019). Similarly, with the utilitarian mandatory ethics setting, each AV makes decisions based on the greater good, prioritizing safety and minimizing harm to individuals and society as a

whole. This aligns with the concept of individual rationality in Nash equilibrium, as each vehicle acts in a way that benefits the overall system.

In Nash equilibrium, no participant has an incentive to unilaterally deviate from their chosen strategy because doing so would not result in a better outcome for themselves. Similarly, if the utilitarian ethical setting was a MES, individual AVs would not have an incentive to deviate from prioritizing the greater good. If a vehicle were to prioritize its own interests at the expense of others, it could potentially lead to negative outcomes such as accidents or inefficient traffic flow, which would not be beneficial for anyone (Schwarting et al., 2019). In this way, the utilitarian MES could provide stability and consistency in the decision-making process of AVs. By adopting a uniform ethical framework that prioritizes the greater good, it establishes a stable and predictable system in which vehicles make consistent choices that optimize overall welfare.

The Nash equilibrium assumes players will make the best choice leading to most optimal outcomes (Schwarting et al., 2019). As seen in the prisoner's dilemma and the "game of chicken" this is obviously not always the case. According to this theory, strategies currently perceived as optimal – such as economic, environmental, corporate, and civil incentives – will not be changed by any players unless given sufficient reason to, such as a player's changed strategy or changed rules of the game.

Application of Nash equilibrium to the utilitarian MES assumes rational decision-making and complete information about the strategies chosen by other vehicles. However, real-world traffic scenarios may involve uncertainties, incomplete information, and varying preferences. Thus, achieving a perfect Nash equilibrium may not always be feasible or desirable in practice. Nonetheless, the concept of Nash equilibrium provides insights into the stability and rationality of the utilitarian approach in the context of choosing an appropriate ethical setting that is optimal for all parties (Schwarting et al., 2019). The Nash equilibrium has applicability to the dilemma of an AVs ethical settings such as civil aspects, economic and corporate incentives to adopt the utilitarian MES and will be referenced in the Civics Lens in Chapter 7.

### 6.1.4 Game Theory Conclusion

In the AV prisoner's dilemma, the dilemma arises from the conflict between individual rationality and the collective interest. Each driver faces the risk of a severe loss if they both defect, even though they would collectively be better off if they both cooperated. In the "game of chicken," the risk lies in the potential harm or damage caused by a collision if both parties refuse to swerve (Acevedo & Krueger, 2005). The outcome can range from a mutual compromise (both parties swerve) to a catastrophic collision if neither party swerves.

Individuals and AVs alike rely on complex decision-making algorithms to navigate various scenarios. The utilitarian ethical setting's algorithms are designed to optimize safety, efficiency, and overall welfare. Through the lens of the prisoner's dilemma and the "game of chicken", AVs should be programmed with the utilitarian MES in order to promote cooperation, minimize harm, and attain optimal outcomes. (Acevedo & Krueger, 2005)

This thesis claims that adopting the utilitarian MES can be equated to the outcome of mutual cooperation in the prisoner's dilemma; the more self-interested PESs - like the egoist PES - can be equated to defection. By programming AVs with a utilitarian MES, the vehicles would prioritize the overall welfare of society, aiming to minimize harm and maximize utility. When all AVs act in the interest of the collective well-being, it can lead to better outcomes for everyone involved. (Acevedo & Krueger, 2005)

By programming AVs with a utilitarian MES it is possible to maximize mutual benefit. In the Prisoner's Dilemma, both players achieve the best outcome when they cooperate. Similarly, in the context of AVs, when vehicles prioritize the greater good and act in a cooperative manner, it can lead to improved traffic flow, reduced congestion, and enhanced safety for all road users (Gogoll & Müller, 2017). This mutually beneficial outcome aligns with the principles of utilitarianism.

By programming AVs with a utilitarian MES it is possible to avoid suboptimal results. If each AV were to act in its own self-interest, similar to betraying in the Prisoner's Dilemma, it could result in suboptimal outcomes. For instance, if every vehicle prioritizes protecting its occupants at the expense of others, it may lead to inefficient traffic patterns, increased congestion, and a higher likelihood of accidents (Gogoll & Müller, 2017). Utilitarian ethics, on the other hand, prioritize the best overall outcome, mitigating the negative consequences that may arise from self-interested decision-making.

By programming AVs with a utilitarian MES it aids in building trust and cooperation. In the Prisoner's Dilemma, players often learn through repeated interactions that cooperation leads to better outcomes in the long run. Similarly, by adopting a utilitarian MES for AVs, it can foster trust and cooperation among different vehicles on the road. This can lead to improved coordination, better adherence to traffic rules, and an overall safer and more efficient transportation system. (Gogoll & Müller, 2017)

While game theory decision-making theorems such as the prisoner's dilemma, "game of chicken", and nash equilibrium provide useful insight into the conceptual framework to understand the potential advantages of a cooperative and utilitarian approach They also provide frameworks for understanding strategic decision-making and ethical considerations in AVs. (Gogoll & Müller, 2017)

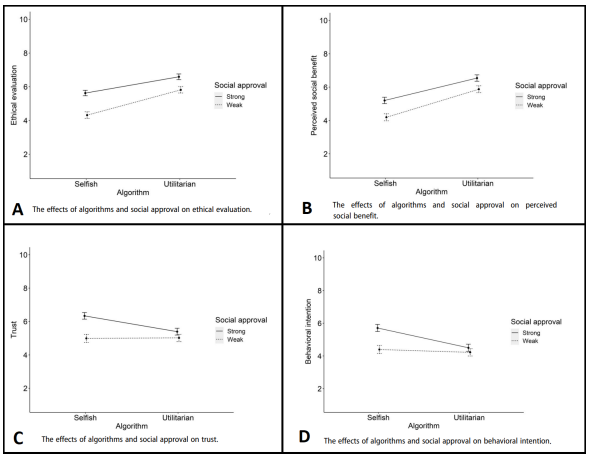
### *6.2 Buyer's Mentality & WTP*

In Section 6.1 we examined an individual's strategic tendencies. In this section, we will examine the behavioral attitudes and intentions of potential AV customers from an individual and social perspective. A customer's decision-making processes leading to purchase, i.e. buyer's mentality, are influenced by many factors, such as: safety issues, liability issues, a customer's WTP, social acceptance and approval, and public trust in AV technology itself and the industry behind it. (Wertenbroch & Skiera, 2002; Liu & Liu, 2021; Joo & Kim, 2022)

In recent studies, it has been proven that a customer's WTP for an AV is inextricably linked to the levels of social approval and the public's trust in AVs and their ethical settings (Joo & Kim, 2022; Liu & Liu, 2021). In their study, Joo and Kim proved that individuals can have conflicting attitudes and behavioral intentions towards AVs. These differing preferences are dependent on whether the issue is an aspect of one's private individual intentions and behaviors or an aspect of those of the general public (Joo & Kim, 2022). This similar dualistic nature was also demonstrated in Section 6.1 and the Ethics Lens in Chapter 5.

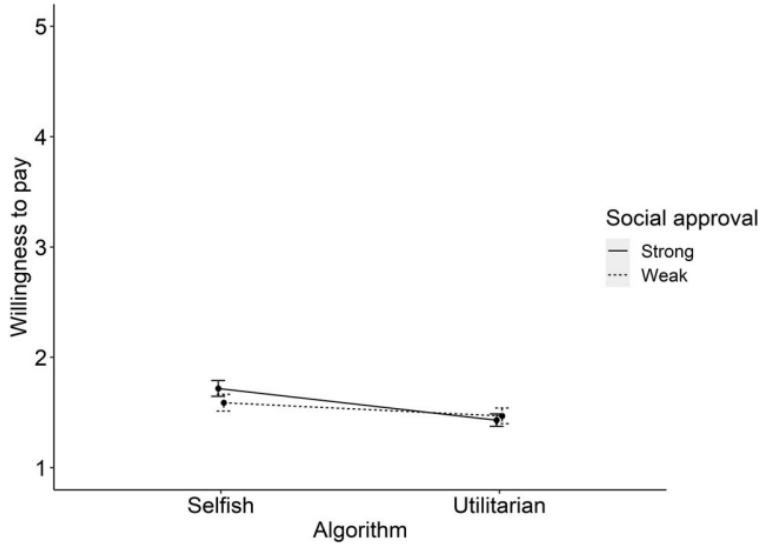
An example of the public having preference that simultaneously differs from their private preferences is the levels of the public's ethical evaluation of an AV's ethical setting compared to the level of an individual's private trust in and behavioral intention to use or purchase AVs (Joo & Kim, 2022). The public's ethical evaluation was equated to the ethical setting's perceived benefits to society as a whole. (Joo & Kim, 2022).

As can be seen in Graph 2, the levels of social approval for the selfish ethical setting and the utilitarian ethical setting are dependent on four different variables: ethical evaluation (A), perceived social benefit (B), trust (C), behavioral intention (D). (Joo & Kim, 2022)




Graph 2. Social approval levels of the selfish and utilitarian ethical settings. (Joo & Kim, 2022)

Positively influencing a customer's behavioral intention to use and WTP for AVs revolves around ethical and consequential considerations, building trust and brand reputation (Joo & Kim, 2022). By adopting the utilitarian MES, AV manufacturers can establish a reputation for prioritizing safety and overall welfare. This can create a positive buyer's mentality, where customers perceive AVs with utilitarian ethics as trustworthy and reliable, leading to increased social acceptance and approval and a WTP for these AVs. Additionally, financial issues like increased insurance costs and safety issues like collision probabilities related to the egoist and utilitarian ethical settings affect the social approval, public trust, and the WTP for AVs. (Contissa et al., 2017;Liu & Liu, 2021; Joo & Kim, 2022)




Graph 3. The effects of the selfish and utilitarian ethical settings on WTP (Joo & Kim, 2022)

AV manufacturers must comply with legal and regulatory frameworks to ensure their vehicles can operate on public roads. In many jurisdictions, safety and public welfare are key considerations in these regulations. Utilitarian ethics align with these concerns, making AVs programmed with the utilitarian MES more likely to meet regulatory requirements. According to Graph 3 and 4, buyers have a preference for AVs that are socially accepted and fully compliant, and this could positively impact their WTP. (Liu & Liu, 2021; Joo & Kim, 2022)

The utilitarian ethical setting prioritizes minimizing harm and maximizing overall welfare, which includes prioritizing safety for everyone. This aspect of utilitarianism appeals to a very broad customer base, as it aligns with the general sentiment of prioritizing the greater good. As safety is a crucial factor for buyers when considering AVs, buyers are likely to perceive AVs programmed with utilitarian ethics as a safer option, making them more inclined to choose these vehicles over alternatives. (ref) Buyers who value

fairness, equity, and the well-being of society as a whole may be more willing to pay for AVs that embrace utilitarian ethics. (Liu & Liu, 2021; Joo & Kim, 2022)

For example, AVs programmed with utilitarian ethics have the potential to reduce accidents and associated costs, such as medical expenses, property damage, and legal fees. Buyers considering the long-term financial implications of owning an AV may see the utilitarian ethical setting as a financially viable choice. The reduced risk of accidents can result in lower insurance premiums and decreased personal liability, making AVs more attractive from a buyer's perspective. (Contissa et al., 2017)

Buyers' mentality often revolves around trust and reputation. By adopting the utilitarian ethical setting, AV manufacturers can establish a reputation for prioritizing safety and overall welfare. This can create a positive buyer's mentality, where customers perceive AVs with utilitarian ethics as trustworthy and reliable, leading to increased WTP for these vehicles. (Liu & Liu, 2021; Joo & Kim, 2022)

AV manufacturers must comply with legal and regulatory frameworks to ensure their vehicles can operate on public roads. In many jurisdictions, safety and public welfare are key considerations in these regulations. Utilitarian ethics align with these concerns, making AVs programmed with utilitarian ethics more likely to meet regulatory requirements. Buyers may have a preference for AVs that are fully compliant, and this could positively impact their WTP. (Liu & Liu, 2021; Joo & Kim, 2022)

Buyer's mentality and WTP can vary among individuals, and some buyers may prioritize factors other than utilitarian ethics, such as personal convenience or customization options. Moreover, there may be ethical concerns associated with the utilitarian ethical setting, such as potential sacrifices of individual autonomy or perceived unfairness. Therefore, the perceived viability of the utilitarian ethical setting should be evaluated alongside other ethical frameworks and individual preferences to ensure a comprehensive understanding of buyer behavior and WTP in the context of AVs. (Liu & Liu, 2021; Joo & Kim, 2022)

### *6.3 Public Acceptance & Trust*

The utilitarian ethical setting can be perceived as the most potentially viable ethical setting for universal application when considering societal aspects such as public acceptance and trust (Liu & Liu, 2021). For example, the utilitarian ethical setting in AVs prioritizes safety and overall welfare. Social acceptance and public trust in AV technology can be enhanced when people perceive that AVs are designed to prioritize their safety and well-being. Knowing that AVs are programmed to make decisions that minimize harm and maximize overall welfare can help instill confidence in their capabilities. (Liu & Liu, 2021; Joo & Kim, 2022)

Public acceptance can be fostered when the ethical setting and decision-making algorithms are transparently communicated to the public. When people understand the underlying principles guiding AV behavior, they are more likely to trust and accept the technology (Liu & Liu, 2021). Predictability in AV actions also contributes to public acceptance, as it allows people to anticipate and understand how AVs will behave in different situations. (Liu & Liu, 2021; Joo & Kim, 2022)

The utilitarian ethical setting in AVs is based on the premise of maximizing overall welfare. Public acceptance can be reinforced by demonstrating the societal benefits of AVs programmed with utilitarian ethics (Liu & Liu, 2021; Joo & Kim, 2022). These benefits can include reducing accidents and injuries caused by human error, enhancing traffic efficiency and reducing congestion, improving transportation accessibility, and reducing environmental impacts. Public awareness of these potential advantages can foster acceptance and support for AV technology.

Public acceptance can be bolstered by involving the public in decision-making processes related to AV deployment and ethical settings (Liu & Liu, 2021). By engaging with individuals, communities, and stakeholders, AV manufacturers and policymakers can solicit input and address concerns regarding ethical considerations. Involving the public in shaping the ethical settings of AVs helps ensure that diverse perspectives and values are taken into account, fostering acceptance and trust. (Liu & Liu, 2021; Joo & Kim, 2022)

Public acceptance can be strengthened by the presence of effective ethical oversight and regulation of AV technology. Robust governance frameworks that incorporate ethical guidelines and standards can help build public trust (Liu & Liu, 2021). Clear regulations regarding the ethical setting of AVs, liability allocation, data privacy, and cybersecurity can address public concerns and provide assurances about the responsible development and deployment of AV technology. (Liu & Liu, 2021; Joo & Kim, 2022)

Public acceptance can be facilitated through education and awareness campaigns that promote understanding of AV technology and its ethical considerations. Transparency, such as providing accessible information about how AVs are programmed, their safety features, and the ethical frameworks guiding their behavior can help alleviate concerns and misconceptions. Increased public knowledge can lead to greater acceptance and trust in AV technology. (Liu & Liu, 2021; Joo & Kim, 2022)

While the utilitarian setting in AVs can contribute to public acceptance and trust, it's important to address and engage with the diverse perspectives and values within society. The utilitarian setting can raise concerns about the perceived fairness and individual autonomy. Some potential buyers may be hesitant to embrace AVs with a utilitarian MES due to the perceived risk of being sacrificed for the greater good. The idea that AVs may make decisions that prioritize overall welfare over individual safety can lead to public distrust and

resistance to adopting this technology. Striking a balance between overall welfare and individual values, privacy concerns, and other ethical considerations is crucial for fostering widespread acceptance and trust in AV technology. (Liu & Liu, 2021; Joo & Kim, 2022)

### *6.4 Behavioral Lens Conclusion*

When considering a customers' WTP, AVs with the utilitarian MES are perceived the most attractive option for universal application for many reasons (Joo & Kim, 2022). For example, AVs programmed with utilitarian ethics prioritize overall welfare and make decisions that prioritize safety and risk reduction for all road users. Customers are likely to value and be willing to pay for AVs that prioritize their safety and reduce the risk of accidents caused by human error. The utilitarian ethical setting can provide a strong selling point for AV manufacturers by emphasizing the enhanced safety features and reduced risk associated with their vehicles.

The utilitarian ethical setting in AVs contributes to the greater good and societal well-being by minimizing overall harm and maximizing overall welfare. This can appeal to customers who value contributing to a safer and more efficient transportation system. Customers who prioritize societal impact and are willing to pay for technologies that align with their values may be more inclined to choose AVs that adhere to utilitarian ethics. (Joo & Kim, 2022)

AV manufacturers that prioritize utilitarian ethics demonstrate a commitment to ethical decision-making and public safety. This commitment can build trust and enhance their reputation in the market. Customers are more likely to trust and choose AVs from manufacturers that prioritize ethical considerations and demonstrate a commitment to the greater good. This can positively impact their WTP a higher price for AVs that prioritize utilitarian ethics.

While the initial cost of AVs may be higher, the long-term cost savings associated with reduced accidents and improved efficiency can make them financially appealing to customers. AVs programmed with utilitarian ethics are designed to minimize the likelihood of accidents caused by human error. This can result in reduced insurance premiums, lower repair costs, and potential savings on medical expenses related to accidents. Customers who perceive AVs as cost-effective and financially advantageous may be more willing to pay for vehicles that prioritize utilitarian ethics.

AV manufacturers that adopt the utilitarian ethical setting may gain a competitive advantage in the market. As the demand for AVs increases, customers are likely to compare different models based on safety, reliability, and ethical considerations. Manufacturers that prioritize utilitarian ethics can differentiate

themselves from competitors by highlighting their commitment to safety and societal well-being. This differentiation can attract customers and give them a competitive edge in the market.

AVs programmed with the utilitarian ethics setting can incorporate advanced safety features that prioritize the overall welfare of all road users. These features may include sophisticated sensors, cameras, and algorithms that enable better detection and response to potential hazards. Customers are likely to perceive the increased safety and reduced risk of accidents as valuable and may be willing to pay a higher price for AVs equipped with these advanced safety technologies.

Utilitarian ethics can be integrated into AVs to improve the overall driving experience. AVs can be programmed to optimize traffic flow, reduce congestion, and minimize travel time. Customers who value convenience and efficiency may be willing to pay a premium for AVs that prioritize utilitarian ethics and offer a smoother and more efficient driving experience.

Utilitarian ethics can be combined with customizable features in AVs. Customers may have the option to adjust the ethical settings within certain boundaries to align with their personal preferences. For example, customers could have the ability to adjust the threshold for avoiding harm to occupants versus pedestrians. This balance between utilitarian ethics and personalization can cater to individual values and increase customers' WTP for AVs that align with their specific ethical preferences.

AVs programmed with utilitarian ethics contribute to the greater good and societal well-being. Customers who value making a positive impact on society may be more willing to pay a premium for AVs that align with their ethical values and promote the well-being of all road users. This aligns with the growing trend of ethical consumerism, where customers are willing to support products and services that prioritize social and ethical considerations.

AV manufacturers that prioritize utilitarian ethics can build a reputation for ethical decision-making and public safety. This reputation can enhance customer trust and increase their WTP for AVs from these manufacturers. Customers who prioritize safety and ethical considerations are more likely to choose AVs from reputable manufacturers, even if they come with a higher price tag.

Another outcome of the utilitarian ethical decision-making setting in AVs could be that when forced to choose between colliding with two different vehicles, an AV would choose to collide with the safer vehicle (Goodall, 2014). Because of this, drivers who pay more for safety may be at greater risk of colliding with another vehicle (Goodall, 2014). Therefore, the utilitarian ethical setting may not be the preferred ethical setting for self-interested individuals or people with family or kin to protect. AV manufacturers also may

reject the utilitarian ethical setting because it may open them up to litigation from vehicle owners. (Goodall, 2014; Smith et al., 2008; Kumfer & Burgess, 2015)

Because of a potential customer's WTP if affected by the AVs ethical system, utilitarianism may not be the preferred ethical setting chosen by AV manufacturers. If a manufacturer chooses a MES separate from the rest of the industry, they could drive away potential customers and allow themselves to be legally liable and open to litigation. (Smith et al., 2008; Kumfer & Burgess, 2015) For this reason as well, the ethical setting should be a mandated industry-wide standard in AVs.

In summary, the utilitarian ethical setting in AVs can be perceived as the most viable ethical setting for universal application when considering customers' WTP. The enhanced safety features, positive societal impact, reputation and trust, long-term cost savings, and competitive advantage associated with AVs programmed with utilitarian ethics can attract customers and justify their WTP for a premium for these vehicles. Nonetheless, it is important for manufacturers to effectively communicate and educate customers about the ethical considerations and benefits of AVs programmed with utilitarian ethics to maximize their market appeal.

It is important to note that these examples are hypothetical and intended to illustrate how the utilitarian ethical setting could potentially increase a customer's WTP for AVs. The specific impact on WTP can vary depending on individual preferences, market dynamics, and the overall value proposition of AVs in terms of safety, convenience, and efficiency.

The utilitarian MES can be seen as favorable from a societal perspective, particularly in terms of public acceptance and trust. If AVs are programmed to prioritize the greater good, it may enhance public trust in this emerging technology. By prioritizing human safety and minimizing the potential for harm, this approach can help overcome ethical concerns associated with AVs and gain societal acceptance. (Liu & Liu, 2021; Joo & Kim, 2022)

Additionally, a standardized ethical setting can help simplify insurance policies and liability frameworks, making it easier to allocate responsibility in the event of accidents or incidents involving AVs. It's important to note that the utilitarian MES is just one ethical approach among many, and it may not be universally accepted or without its criticisms (Liu & Liu, 2021; Joo & Kim, 2022). Different perspectives and ethical frameworks can lead to different conclusions, and the choice of an ethical setting for AVs requires careful consideration, incorporating various stakeholders' input and ethical deliberation.

These perspectives provide additional interdisciplinary viewpoints to consider the utilitarian MES for AVs as the best ethical setting. It's important to note that ethical considerations should involve a comprehensive

analysis of multiple perspectives and stakeholders' input to ensure a well-rounded approach to the ethical decision-making processes of AVs.

## 7 Civics Lens

Through the last lenses in Chapter 5 and 6, we examined the theoretical and practical ethical aspects of individuals. We examined how people ideally ought to act according to the relevant normative ethical theories, and how individuals actually behave according to behavioral and psychological theories. In this chapter, through this lens, we'll examine the more social and civil aspects from a perspective of optimal group behavior when taking group dynamics and certain facts about human nature into consideration.

Civics deals with a citizen's rights, obligations, and the study of the interrelationship between individual citizens and the rest of society. The study of civics is also interdisciplinary as it includes the study of civil codes and the people's role in maintaining proper governmental operation and accountability (Plato, 1963). How can we civically optimize the organism/environment relationship between individuals and the whole? How can we optimize the balance, not just between ourselves and the rest of society, but also within the environment we share? How can we find harmonious equilibrium and form the most optimal civilization? This lens will attempt to answer those questions within the realm of AVs.

AVs have the potential to revolutionize transportation and change the landscape of society and civilization as we know it, but the widespread adoption of AVs presents complex ethical and behavioral considerations that must be addressed to ensure their responsible integration into civil society (Emory et al., 2022). The utilitarian ethical setting, aims to maximize overall societal welfare. This lens focuses its examination on the interconnected societal and civil aspects of adopting the mandatory utilitarian ethical setting in AVs, including engineering and safety, environmental stability, economic savings, corporate and public policy incentives, and the importance of coming to a social consensus.

### *7.1 Engineering & Safety*

AVs require engineering decisions that prioritize minimizing harm and maximizing overall safety. This involves developing robust collision-avoidance systems, advanced sensor technologies, and reliable algorithms for decision-making in complex scenarios. By focusing on ensuring passenger and pedestrian

safety, reducing accidents, and minimizing overall harm to all stakeholders in traffic situations, utilitarian based engineering decisions play a crucial role in AVs. (Gogoll & Müller, 2017)

Research indicates that AVs equipped with advanced sensor technologies have shown a 90% reduction in accidents compared to traditional vehicles (Litman, 2014). It can be inferred that the driver's human error is the leading cause of motor vehicle collisions (Litman, 2014). Using the logical rule of transference then, driverless AVs are multitudes safer than the current 'horseless carriages'.

AVs are utilizing cutting-edge sensor technologies, such as LiDAR and radar systems, enabling AVs to detect and respond to their surroundings with exceptional accuracy. Advanced collision-avoidance systems and reliable decision-making algorithms are engineered to mitigate the risk of accidents, ensuring passenger and pedestrian safety. AVs can optimize safety, reduce accidents caused by human errors, ensure consistency and reliability, proactively mitigate risks, and continuously learn and improve. Simultaneously, advancements in sensor technologies and collision-avoidance systems enhance AV safety, protecting passengers, pedestrians, and the environment. (Morfeldt & Johansson, 2022)

Standardization, interoperability, and predictable behavior can be achieved through the utilitarian ethical setting. Utilizing the utilitarian MES, engineering challenges related to the development and deployment of connected AVs (CAVs) are more straightforward with standards and regulations in place. The transition from traditional transportation to AVs is also less turbulent with a utilitarian MES (Gogoll & Müller, 2017). CAVs could have the possibility to synchronize with each other and have a 'swarm behavior' and traffic could flow more efficiently and at optimal levels much higher than we are currently used to. (Morfeldt & Johansson, 2022; Yu et al., 2021; Ulbrich et al., 2020)

### *7.2 Environment & Sustainability*

AV technology has many benefits, one of them being that AVs can positively impact the environment and contribute to a significant reduction in greenhouse gas emission. By the year 2050, the global adoption of AVs could reduce global greenhouse gas emissions by up to 60% (Morfeldt & Johansson, 2022). By optimizing traffic flow, through various means, such as: more efficient routing algorithms that minimize travel distances and congestion, encouraging shared mobility, ride-sharing services to optimize vehicle utilization, and promoting fuel-efficient driving behaviors, AVs can contribute to environmental sustainability. In this way, AVs can contribute to improving air quality and reducing the negative health impacts associated with pollution.

AVs programmed with utilitarian ethics can prioritize transportation options and accessibility in underserved and marginalized communities, reducing transportation-related inequalities. This can promote equitable

access to sustainable transportation and mitigate the environmental injustices often faced by disadvantaged populations. This can have a positive effect on marginalized communities that are disproportionately affected by environmental hazards from mass transportation and can also contribute to social justice and more equitable access to clean and sustainable transportation options. (Gogoll & Müller, 2017)

With the adoption of AV technology, the transportation of goods and services, and commutes could be revolutionized to the point that individuals no longer need to worry about transportation methods or routes or even own a personal mode of transportation (Morfeldt & Johansson, 2022). The adoption of AVs and shared mobility services can reduce the need for individual car ownership which is connected to reduced personal liability, reduced insurance costs and vehicle maintenance costs for everyone in society (Contissa et al., 2017). By optimizing vehicle utilization and facilitating ride-sharing, AVs can help reduce the number of vehicles on the road, leading to decreased congestion, lower resource consumption, and less environmental impact. This is another way AVs can promote sustainable transportation and reduce the ecological footprint associated with transport, commute, and car ownership. (Morfeldt & Johansson, 2022)

AVs programmed with utilitarian ethics can influence urban planning and land use patterns in a sustainable manner. By optimizing traffic flow and reducing congestion, AVs can help minimize the need for extensive road infrastructure and decrease demand for parking spaces, freeing up space for green areas, public transportation, and pedestrian-friendly environments. This can enhance the quality of urban life and promote sustainable urban development. (Morfeldt & Johansson, 2022)

Optimization of AV technology will most likely lead to the development of electric vehicles transported on less frictional pathways by, presumably, electronic and magnetic forces. This would probably lead to the integration of renewable energy sources. AVs programmed with utilitarian ethics can be integrated with renewable energy sources, such as electric vehicle charging infrastructure powered by renewable sources (Morfeldt & Johansson, 2022). This combination can further enhance the environmental sustainability of AVs by reducing reliance on fossil fuels and promoting the use of clean energy. By aligning with renewable energy goals, AVs can contribute to a more sustainable and eco-friendly transportation system.

The utilitarian ethical setting in AVs can take into account the entire lifecycle of vehicles, including manufacturing, operation, and disposal. By considering the environmental impact at each stage, AVs can be designed and operated in a manner that minimizes resource consumption, waste generation of planned obsolescence, and pollution (Morfeldt & Johansson, 2022). This holistic approach to sustainability aligns with the principles of utilitarian ethics, which seek to maximize overall welfare and minimize harm in a broader sense that includes all biological life in our shared environment.

In these ways, and many more, the utilitarian ethical setting can be seen as best by reducing the negative impact to the environment and promoting sustainability. However, examination of the utilitarian ethical setting in AVs should be accompanied by careful considerations and safeguards to ensure that environmental goals are not pursued at the expense of other ethical considerations. Balancing the needs of the environment with other aspects such as social justice, safety, and fairness requires a comprehensive approach that takes into account multiple ethical frameworks and engages stakeholders from various domains.

### *7.3 Economic & Corporate Incentives*

Corporations and policy makers have incentives to adopt utilitarian ethical settings for AVs due to their potential benefits. For corporations, the implementation of the utilitarian MES can enhance social acceptance and brand reputation and a customer's WTP by prioritizing safety and environmental concerns (Liu & Liu, 2021; Joo & Kim, 2022). Additionally, economic savings resulting from utilitarian engineering can lead to cost reductions and increased profitability. (Othman. 2022)

The Boston Consulting Group estimated that by 2035, AVs could save individuals and businesses approximately $1.1 trillion per year through reduced fuel consumption and improved productivity (Othman. 2022). This is a substantial saving of resources and economic wealth. For these reasons and many others, policymakers may also have an interest in adopting the utilitarian MES for AVs to promote public safety, environmental sustainability, and economic efficiency.

Adopting the utilitarian MES would lead to engineering decisions leading to safer roads and improved collision-avoidance systems. Moreover, the optimized traffic flow mentioned earlier can contribute to economic savings by improving productivity, lowering fuel consumption expenses, and decreasing accident-related expenses to corporations and society, including medical bills, property damages, insurance claims, and vehicle repairs (Contissa et al., 2017). Additionally, the efficient use of resources through shared mobility and ride-sharing services can lead to cost savings for individuals and businesses, reducing the economic burden on individuals, insurance companies, and society as a whole. (Othman. 2022)

A buyer's mentality affects consumer demand and plays a significant role in shaping the market for AVs. Buyers' preferences and WTP may be influenced by factors other than utilitarian ethics, such as luxury features, brand reputation, or individual customization options (Liu & Liu, 2021; Joo & Kim, 2022). If consumers prioritize these factors over utilitarian ethics, AV manufacturers may be incentivized to cater to those preferences to capture market share and maximize profits. (Schreurs & Steuwer, 2016)

AV manufacturers must consider the economic viability of their products. The utilitarian ethical setting may require additional investments in advanced technology, safety features, and extensive testing and validation

processes. These factors can increase the production costs of AVs, potentially impacting their affordability and market competitiveness (Liu & Liu, 2021; Joo & Kim, 2022). Manufacturers may need to strike a balance between the utilitarian ethical setting and cost considerations to ensure the economic viability of their AVs.

Ethical considerations, including the utilitarian ethical setting, can and should be integrated into corporate values and economic strategies. Some companies may choose to prioritize ethics and the greater good as part of their corporate social responsibility initiatives, even if it comes at some economic cost or other initial disadvantage (Liu & Liu, 2021; Joo & Kim, 2022). Due to the capitalistic system, the universal application of the utilitarian ethical setting in AVs may face challenges due to the complex interplay between corporate incentives, consumer demand, and economic considerations.

While utilitarian ethics may align with certain regulatory requirements focused on safety and overall welfare, other policies might prioritize economic growth, job creation, or specific industries. AV manufacturers might be influenced by policy incentives that prioritize certain economic goals, such as maximizing shareholder profit, over strict adherence to utilitarian ethics. Government regulations and policy incentives can shape the economic landscape for AVs and help create a smooth transition for individuals, businesses, and society as a whole. (Gogoll & Müller, 2017)

It is crucial for society to reach a collective and overlapping consensus on ethical guidelines for AVs (Gogoll & Müller, 2017). By involving stakeholders, including engineers, manufacturers, policymakers, ethicists, and the general public, consensus can be reached on key ethical priorities such as preserving human life, preventing unnecessary harm, and ensuring safety, transparency and accountability in decision-making algorithms.

### *7.4 Civil Lens Conclusion*

According to the theory of group dynamics, a group has five sequential stages: forming, storming, norming, performing, and adjourning. This thesis claims, according to the information obtained through these interdisciplinary lenses, that we can norm through collective and overlapping consensus on the ethical settings of AVs. Only then can we start performing at our optimal levels. best, in theory according to the data above, with the utilitarian MES.

Through this lens, there are clear ethical, environmental, safety, economic, and regulatory incentives for corporations and policy makers to adopt utilitarian MES. By considering the potential advantages in terms of social approval, trust, cost reductions, WTP, public safety, sustainability. and various aspects of their

interplay, we can promote consensus, stakeholders can collectively foster the responsible development and widespread adoption of AVs while maximizing safety, efficiency, and sustainability for society as a whole.

## 8 Discussion

According to the Ethical Lens, adopting the utilitarian MES can be seen as optimal from each ethical viewpoint. The utilitarian MES: is best for others – as altruists would prefer, is consistent with virtuous behavior – as virtue ethicists would prefer, is acting as one wishes everyone else would – as Kant's deontological Categorical Imperative would prefer, is beneficial for the majority – as utilitarians would prefer, maximizes one's own best interest – as ethical egoist would prefer.

According to the game theoretical results of the prisoner's dilemma and 'game of chicken', mandating the implementation and deployment of AVs with a utilitarian MES could be considered an ethical duty and obligation we owe ourselves as a civilization, according to the tenets of deontology, virtue ethics, utilitarianism, and altruism. Four out of five of the main normative ethical viewpoints would ethically approve of this line of action, egoism being the obvious exception.

From a behavioral and psychological perspective, the utilitarian MES can be beneficial in promoting trust and predictability among passengers and pedestrians (Joo & Kim, 2022). Individuals often expect others to act in ways that prioritize safety and well-being, and AVs following a utilitarian approach align with these expectations. This setting may also contribute to reducing anxiety and uncertainty associated with AV interactions, leading to more positive user experiences, societal acceptance and approval, and public trust. (Joo & Kim, 2022)

A PES like the Ethical Knob may improve social acceptance and public trust of AVs by enabling users the freedom to choose the ethical setting that best represents their ethical outlooks and moral principles. The Ethical Knob could also help smooth the transition of AVs to the market (Contissa et al., 2017). However, in the debate between the PES and MES in AVs, the utilitarian MES can be seen as the superior option for several reasons:

The utilitarian MES provides a consistent and fair approach to AV decision-making. It treats all individuals equally, prioritizing overall welfare and minimizing harm. In contrast, a PES allows individual AV owners to

define their own ethical settings, which can lead to inconsistencies and varying moral standards. The utilitarian MES ensures a level playing field where all AVs adhere to a common ethical framework.

The utilitarian MES aligns with the principle of maximizing overall welfare. AVs programmed with utilitarian ethics aim to minimize harm and promote the greatest good for the greatest number. This approach prioritizes the well-being and safety of all road users, contributing to a more socially beneficial transportation system compared to individualized PESs, which may prioritize the interests of specific occupants over others.

AVs often encounter challenging moral dilemmas, such as the classic "trolley problem" scenarios. The utilitarian MES offers a clear decision-making framework to navigate these situations by focusing on minimizing overall harm. In contrast, PESs may introduce subjective and potentially biased ethical settings, leading to inconsistent and potentially unfair outcomes.

Implementing a utilitarian MES in AVs can enhance safety and promote public acceptance (Liu & Liu, 2021; Joo & Kim, 2022). By prioritizing overall welfare, AVs programmed with utilitarian ethics can make more rational and objective decisions, reducing accidents caused by human error and improving overall road safety. The transparency and predictability of a universal ethical standard can also help build public trust and confidence in AV technology. (Liu & Liu, 2021; Joo & Kim, 2022)

From a legal and regulatory perspective, the utilitarian MES offers clear guidelines for liability and accountability. It provides a standardized approach that facilitates legal frameworks, ascribing responsibility in unavoidable accidents. In contrast, accommodating various PESs in the legal system can create complexity and challenges in determining liability and allocating accountability.

Embracing the utilitarian MES represents a step toward ethical progress and optimization in AV technology. By implementing an ethical framework that prioritizes overall welfare, AVs can contribute to reducing accidents, saving lives, and enhancing transportation efficiency. This aligns with the broader goal of continuously improving ethical decision-making and creating a safer and more ethically sound transportation system.

While the utilitarian MES offers several advantages, it is important to acknowledge and address concerns related to individual rights, justice, and cultural differences. Striking a balance between utilitarian principles and other ethical considerations, alongside robust legal and regulatory frameworks, can help mitigate these concerns and create a more comprehensive and ethically sound approach to AV technology.

In conclusion, the utilitarian MES can be considered the superior option in the debate between PES and MES in AVs due to its consistency, fairness, focus on overall welfare, ability to address moral dilemmas, enhanced safety, legal clarity, and potential for ethical progress. Nonetheless, ongoing discussions, stakeholder engagement, and multidisciplinary collaboration are crucial to ensure the ethical, legal, and societal implications of the utilitarian MES are thoroughly considered and implemented in a manner that balances the interests of all stakeholders involved.

## 9 Conclusion

As can be seen through the lenses of this thesis, the adoption of the utilitarian MES can have effects that reverberate through many areas of society and aspects of life. Through these lenses, we can see how these separate disciplines and issues are interconnected and interrelated. The adoption of the utilitarian MES as an industry standard could have an exponential effect on the wellbeing of individuals and society alike. The improvement of quality and efficiency of transportation can have a cumulative effect on the environment and overall quality of life. possibly leading to a real paradigm shift in the ways humans are, be, and do.

People want the utilitarian ethical setting to be adopted by the public but privately moved by more morally weak selfish tendencies. They want others to have it. On an ethical level/as ideal. On a society level/for the use of most of society. They also want to adopt it because it is theoretically the safest option with the least casualties. They want it to be used. But they don't want to be a sucker if others can choose more personally advantages options (jealousy and envy are familiar vices)  It seems without proper leadership (player making the first benevolent move that the other player can mirror) or a catalyzing event, we the public and stakeholders are stuck in a suboptimal loop / livelock?

To stay true to human psychology & behaviorism, people will need to walk the path and see the benefits in practice before being completely convinced of the benefits & merits of the virtuous act of being willing to “take one for the team”/sacrifice oneself for the greater good.

## 10 Final Thesis Conclusion

As automation and autonomous processes like AVs become more prevalent, it becomes crucial to determine who should be held accountable for the actions and decisions made by these systems. Ethical questions arise regarding the allocation of responsibility when a machine or algorithm makes a harmful or unethical choice. Resolving these issues requires establishing frameworks for accountability and assigning clear lines of responsibility.

Automated and autonomous processes rely on algorithms to make decisions and predictions. However, these algorithms can be biased, reflecting the prejudices and biases of their creators or the data they were trained on. As a result, issues related to algorithmic bias and discrimination emerge, leading to concerns about fairness and equity. Resolving these ethical dilemmas involves developing unbiased algorithms, ensuring diverse representation during their development, and establishing guidelines for addressing discriminatory outcomes.

Autonomous entities can make decisions that have significant consequences for human life and well-being. This raises ethical questions about the criteria used in decision-making processes, the trade-offs involved, and the level of human oversight required. Resolving these dilemmas entails defining ethical guidelines and regulatory frameworks that govern the behavior and decision-making capabilities of autonomous systems.

The increasing prominence of automatic processes and autonomous entities necessitates a deeper examination of the ethical implications and a concerted effort to address the disputed ethical issues. It requires collaboration between stakeholders from various fields, including technology, ethics, policy-making, and academia, to establish frameworks, guidelines, and regulations that ensure the responsible and ethical development of these autonomous technologies.

Because of the arrival and increased global production of automation, and autonomous processes and entities, seemingly ubiquitous and omnipresent in modern society, there is a need to program these automatic and autonomous processes ethically. This increased urgency can almost be felt like a volcano brewing – and the pressure is rising. If we don’t act now, we may have to react later. Can we come to a unified overlapping consensus and reach answers to these questions before it's too late?

Role Virtues should be adopted by everyone, not just engineers as was mentioned in the virtue ethics section., but individuals in all sorts of different roles and professions. From programmers to policy makers to the general public, we should all be doing our part by living up to our ideal’s potential. As Marcus Aurelius

famously stated, "We are born for cooperation.". According to the AVs prisoner's dilemma, the results of the MATLAB computer simulations, and the MIT Moral Machine statistics, his statement holds true.

Nobody wants to be "that guy" who could be considered responsible or liable for any unfortunate consequences for making the executive decision to adopt the utilitarian MES despite the valid arguments of others. One could imagine feeling hesitant or uncomfortable meddling in another's affairs without their consent, especially if it's against their personal beliefs. Not to mention having to do that on a global level with corporations and captains of industry. Imagine the resistance one would incur trying to convince, not just one CEO or a board of CEOS, but a whole industry. However, there seems to be an urgent ethical need for someone to be the first player to make a benevolent move so that the rest of the players can begin to mirror this positive behavior leading to the most optimal outcome.

Artificial Intelligence and automated processes and autonomous entities could make our government, business management, and daily processes so much more efficient and streamlined and resources allocated with less waste, corruption, and incompetence. This thesis is, in part, meant to be a call to action, at least insofar as to help inspire the development of an ethical testing system or a method to achieve a more informed and unified collective consensus on the ethical dilemmas of autonomous processes.

### *10.1 Paradigm Shift*

A paradigm shift occurs when new ideas or technologies reshape the established conventions, usually making them obsolete in the process, and enable new ways of doing and being. A paradigm shift fundamentally changes how people live. An example of a paradigm shift was the invention and adoption of the automobile.

When the automobile was first introduced and horse-drawn carriages were the norm, New York City, otherwise known as 'The Big Apple', had a population of about 100,000 horses. These horses were necessary for the transport of goods and services. The people of New York thought the new automobiles were a nuisance and a threat (Kavrell, J. 2019). These new automobiles were fast and loud, and they scared the horses.

Simultaneously, The Big Apple was being crippled by 2.5m pounds of manure that these horses produced on a daily basis. This manure problem was so bad that it clogged the streets and was dubbed 'The Big Crapple'. These new 'horseless carriages', while new and loud, were also without the hassles that come with horses, such as: unpredictable animal behavior, the need for stables, rest, provisions, medical care, and carcass and manure removal. (Kavrell, J. 2019)

Within a short period of time, the “horseshit problem of New York” was solved by horseless carriages (Kavrell, J. 2019). Unbeknownst to them at the time, the introduction of the automobile would have unforeseen consequences, many of them positive. Points in time like these, when innovative jumps in advancement disrupt the old paradigm and create a new paradigm, are termed ‘inflection points’.

Mobility has had many great inflection points: the invention of the wheel, husbandry like the domestication of horses and addition of a carriage, the invention of the horseless carriage. and most recently, the invention of the ‘driverless carriage’ – the AV.

One of the goals of this thesis is to get the reader to understand that we are on the cusp of a paradigm shift. Our actions, as individuals, as a society, and as a civilization, will determine our trajectory and degree of inflection into the next paradigm. It is the hope of this author that we choose the right direction at each moral crossroad.

Humans have such technical prowess as to create artificial intelligence, autonomous processes, and quantum supercomputers with machine learning that can analyze big data sets and run incredibly complex simulations (Nielsen & Chuang, 2011). How will we, as a civil society, utilize this new disruptive technology? Will we, as a species, demonstrate wisdom and other virtuous traits alongside our creativity and cunning intelligence? Will we use these great new powers responsibly with ethical integrity and civility – or are we still in our moral and ‘technological adolescence’? (Lemarchand, 2006)

---

Afterword: In this thesis, by including the call-to-action and message of ethical urgency to achieve a collective and overlapping consensus, I’m doing my civic and ethical duty, as a citizen and a scholar, to do what I can for the greater good by using the mediums, platforms, and tools at my disposal, even if it means a lower academic appreciation of this thesis. I think this is an example of consistently living as a virtuous agent, not only when it is convenient. Virtues are strived for, earned, and cultivated. Being virtuous is an active process, and we need to consider these issues seriously and actively come to a collective consensus and change the world for the better.

**References**


Acevedo, M., & Krueger, J. I. (2005). Evidential Reasoning in the Prisoner's Dilemma. The American Journal of Psychology, 118(3), 431–457. http://www.jstor.org/stable/30039074

Aumann, R. (1985). "What is Game Theory Trying to accomplish?" (PDF). In Arrow, K.; Honkapohja, S. (eds.). Frontiers of Economics. Oxford: Basil Blackwell. pp. 909–924.

Awad, E., Dsouza, S., Kim, R. et al. (2018) The Moral Machine experiment. Nature 563, 59–64. https://doi.org/10.1038/s41586-018-0637-6

Ben-Zeev, A. (1982). Who Is a Rational Agent? Canadian Journal of Philosophy, 12(4), 647–661. http://www.jstor.org/stable/40231284

Bloomfield, M. (1993) The Automated Society; A View of the Distant Past, the Present and the Far Future. Masefield Books. Retrieved 29 March 2017

Brueckner, Martin. (2013). Corporation as Psychopath. doi: 10.1007/978-3-642-28036-8_128.

Burkley, E., Burkley, M., Curtis, J., Hatvany, T. (2018). Lead us not into temptation: The seven deadly sins as a taxonomy of temptations. Social and Personality Psychology Compass. 12. 10.1111/spc3.12416.

Böhme, G. & Stehr, N. (2012) The Knowledge Society: The Growing Impact of Scientific Knowledge on Social Relations. Springer Science & Business Media. Retrieved 27 March 2017

Cairns, P., & Cox, A. (Eds.). (2008). Research Methods for Human-Computer Interaction. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511814570

Chang, D. (2008) Comparison of Crash Fatalities by Sex and Age Group.NHTSA, U.S. Department of Transportation, Washington, D.C.

Columbus, P. J., & Rice, D. L. (2017). Alan Watts—In the academy: Essays and lectures. State University of New York Press.

Contissa, G., Lagioia, F. & Sartor, G. (2017) The Ethical Knob: ethically-customisable automated vehicles and the law. Artif Intell Law 25, 365–378. https://doi.org/10.1007/s10506-017-9211-z

D'Souza, J., Burnham, K.J., Pickering, J.E. (2022) Modelling and Simulation of an Autonomous Vehicle Ethical Steering Control System (ESCS) 26th International Conference on Methods and Models in Automation and Robotics, MMAR 2022 - Proceedings pp. 76-80

Emory, K., Douma, F., Cao, J. (2022). Autonomous vehicle policies with equity implications: Patterns and gaps. Transportation Research Interdisciplinary Perspectives. 13. 100521. 10.1016/j.trip.2021.100521.

Geiger, I. (2011). Rational Feelings and Moral Agency. Kantian Review, 16(2), 283-308. doi:10.1017/S1369415410000038

Goodall, N. J. (2014). Ethical Decision Making during Automated Vehicle Crashes. Transportation Research Record, 2424(1), 58–65. https://doi.org/10.3141/2424-07

Hart, W. D. (1996). "Dualism." pp. 265–267 in A Companion to the Philosophy of Mind, edited by S. Guttenplan. Oxford: Blackwell.

Hursthouse, R. & Pettigrove, G. (2022) "Virtue Ethics", The Stanford Encyclopedia of Philosophy, Edward N. Zalta & Uri Nodelman (eds.), Visited 01.05.2023.

Joo, Y. K. & Kim, B. (2022). Selfish but Socially Approved: The Effects of Perceived Collision Algorithms and Social Approval on Attitudes toward Autonomous Vehicles, International Journal of Human–Computer Interaction, doi: 10.1080/10447318.2022.2102716

Kant, I. (2003). Critique of pure reason (M. Weigelt, Trans.). Penguin Classics.

Kavrell, J. (2019). The Future of the World is Going to be Driven by Autonomous Cars (Bachelor's thesis, University of Arizona, Tucson, USA).

King James Bible. (2017). King James Bible Online. https://www.kingjamesbibleonline.org/ (Original work published 1769)

Lee, Y. Yang, H., Wang, M. (2009). Daoist Harmony as a Chinese Philosophy and Psychology. Peace and Conflict Studies. 16. 68-81. 10.46743/1082-7307/2009.1103.

Lemarchand, G. (2006). A Description of the Terrestrial Technological Adolescence Case.) Book: The future of life and the future of our civilization (pp.457-468) Chapter: The Life-time of technological civilizations Publisher: Springer

Litman, T. (2014) Autonomous Vehicle Implementation Predictions: Implications for Transport Planning. Victoria Transport Policy Institute, Victoria, British Columbia, Canada.

Liu, P. & Liu, J. (2021). Selfish or Utilitarian Automated Vehicles? Deontological Evaluation and Public Acceptance, International Journal of Human–Computer Interaction, 37:13, 1231-1242, DOI: 10.1080/10447318.2021.1876357

MacAskill, W. ( 2017). "Effective Altruism: Introduction". Essays in Philosophy. 18 (1): 2. doi:10.7710/1526-0569.1580.

Mill, J. S. (1863) Utilitarianism. London, Parker, son, and Bourn. [Pdf] Retrieved from the Library of Congress, https://www.loc.gov/item/11015966/.

Miller, T. C. (1993). The Duality of Human Nature. Politics and the Life Sciences, 12(2), 221–241. http://www.jstor.org/stable/4235959

Morfeldt, J., & Johansson, D. J. A. (2022). Impacts of shared mobility on vehicle lifetimes and on the carbon footprint of electric vehicles. Nature communications, 13(1), 6400. https://doi.org/10.1038/s41467-022-33666-2

Nielsen, M. A., & Chuang, I. L. (2011). Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press.

Nyholm, S., Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: an applied trolley problem? Ethical theory and moral practice pp 1–15, doi:10.1007/s10677-016-9745-2

Okasha, S. (2011) "Biological Altruism". Stanford Encyclopedia of Philosophy. Visited 13 May 2011.

Othman K. (2022). Exploring the implications of autonomous vehicles: a comprehensive review. Innovative Infrastructure Solutions, 7(2), 165. https://doi.org/10.1007/s41062-022-00763-6

Pigou, A. (2002). The Economics of Welfare (1st ed.). Routledge. https://doi.org/10.4324/9781351304368

Plato. (1963). The trial and death of Socrates; Euthyphro, Apology, Crito, Phaedo. London : New York :Dent; Dutton,

Pope, A. (1963). Butt, J. (ed.). The Poems of Alexander Pope (a one-volume edition of the Twickenham text ed.). Yale University Press.

Poundstone, W. (1993). Prisoner’s dilemma. Anchor. Rapoport, A. (1967). A note on the “index of cooperation” for prisoner’s dilemma. The Journal of Conflict Resolution, 11, 100–103.

Rosebury, B. (2021). Informed Altruism and Utilitarianism. Social Theory and Practice. 47. 717-746. 10.5840/soctheorpract2021922140.

Sanders, S. M. (1988) Is egoism morally defensible? Philosophia 18 (2-3):191-209 (1988) https://doi.org/10.1007/BF02380076

Scheffler, S. (1988). Consequentialism and Its Critics. Oxford: Oxford University Press. ISBN 978-0-19-875073-4.

Schreurs, M. & Steuwer, S. (2016). Autonomous Driving—Political, Legal, Social, and Sustainability Dimensions. 10.1007/978-3-662-48847-8_8.

Schwarting, W., Pierson, A., Alonso-Mora, J., Karaman, S., & Rus, D. (2019). Social behavior for autonomous vehicles. Proceedings of the National Academy of Sciences of the United States of America, 116(50), 24972–24978. https://doi.org/10.1073/pnas.1820676116

Sell, D. (2023). The ethical reasoning and programming logic behind the ethical decision-making processes of autonomous vehicles. Bachelor's thesis, Tampere University. https://urn.fi/URN:NBN:fi:tuni-202306026470

Shaver, R. (2019), "Egoism", in Zalta, Edward N. (ed.), The Stanford Encyclopedia of Philosophy (Spring 2019 ed.), Metaphysics Research Lab, Stanford University, retrieved 2020-05-27

Smith, J. H., P. M. Harper, and R. A. Burgess. (2008) Engineering Ethics: Con-cepts, Viewpoints, Cases, and Codes. National Institute for Engineering Ethics, Lubbock, Tex.

Stevenson, C. L. (1937). "The Emotive Meaning of Ethical Terms". In Stevenson, C. L. (ed.). Facts and Values. Yale University Press (published 1963).

Ulbrich, F., Rotter, S., Rojas, R. (2020). Adapting to the Traffic Swarm: Swarm Behaviour for Autonomous Cars. 10.4018/978-1-7998-1754-3.ch067.

Waller, B. N. (2005) Consider Ethics: Theory, Readings, and Contemporary Issues. New York: Pearson Longman: 23

Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. Journal of Marketing Research, 39(2), 228–242. https://doi.org/10.1509/jmkr.39.2.228.19086

Yu, H., Jiang, R., He, Z., Zheng, Z., Li, L., Liu, R., Chen, X. (2021). Automated Vehicle-involved Traffic Flow Studies: A Survey of Assumptions, Models, Speculations, and Perspectives. Transportation Research Part C Emerging Technologies. 10.1016/j.trc.2021.103101.

Zahavi, A. (1995). Altruism as a Handicap: The Limitations of Kin Selection and Reciprocity. Journal of Avian Biology, 26(1), 1–3. https://doi.org/10.2307/3677205