Tampere University

Daniel Sell

THE ETHICAL REASONING AND PROGRAMMING LOGIC BEHIND THE ETHICAL DECISION-MAKING PROCESSES OF AUTONOMOUS VEHICLES

Faculty of Information Technology and Communication Sciences
Bachelor's Thesis
May 2023

ABSTRACT

Daniel Sell: The ethical reasoning and programming logic behind the ethical decision-making processes of autonomous vehicles
Bachelor's Thesis
Tampere University
Bachelor's Degree Programme in Computer Sciences
May 2023

In this thesis, the foundations behind the ethical reasoning and logical programming of the ethical decision-making processes of autonomous vehicles (AVs) in an unavoidable crash scenario will be examined through the interdisciplinary lenses of philosophy (ethics) and computer science (programming).

The technoethical issues of AVs and their ethical decision-making processes will be examined from various different ethical standpoints. These ethical stances will then be further delineated through corresponding pseudocode in the programming language of Python.

It is the goal of this thesis to analyze these ethical concepts clearly and comprehensively, in a multidisciplinary manner, and represent their contemporary significance and relevance in the realm of autonomous technologies. These central concepts need to be considered and understood by the general public and stakeholders in order for them to be able to participate in the greater debate surrounding these issues. In this way, as an advanced civilization, we can finally come to a general consensus on these ethical problems and apply appropriate solutions in the real world.

Keywords: ethical decision-making, autonomous vehicles (AVs), deontology, utilitarianism, virtue ethics, egoism, altruism, programming, pseudocode

TABLE OF CONTENTS

1 Introduction

The world is ever increasingly going technological, digital, virtual, online. Many of our everyday processes are becoming automated and autonomous, from robotic assembly lines to autonomous vehicles (AVs). Students and educators from all fields should be aware of how this technological evolution affects their fields and their lives. Society as a whole should be aware of the implications of the evolution of these technological advances to be able to adapt and evolve with them.

Specifically, the practitioners of the computer sciences should be especially mindful of the ethical consequences and causal ramifications of their work. Coders should learn not just how to code working programs, but how to program responsibly, with accountability and ethical integrity. Technology is having an ever increasing affect on our lives, ethics and technology are increasingly becoming intertwined. This technology needs to be designed and made in an intelligent and wise manner, implementing 'practical wisdom', with ethical integrity. (Smith et al., 2008)

Recent developments in autonomous technologies, have caused a rising urgency to address ethical issues, which have hitherto been contentiously debated or left open-ended for lack of proof and/or consensus. The question then becomes, "How can we program our ethical preferences into autonomous processes without knowing ourselves what our own collective preferences are?" Therefore, before we can ethically autonomize processes on a global scale, we need to come to a general consensus on a collective ethical standpoint, as it affects everyone, all biological life even, worldwide. These increasingly urgent moral dilemmas are sufficiently significant and contemporarily relevant enough to warrant closer examination (Kumfer & Burgess, 2015). These dilemmas need solutions and we need to come to a consensus as an advanced civilization.

To address a tiny sliver of the greater contemporary ethical issues of this subject, this thesis will focus on understanding the ethical reasoning and programming logic behind an AV's ethical decision-making processes in an unavoidable crash scenario. This issue was chosen as the focus of this thesis because it clearly demonstrates the consequential connection between ethics and autonomous processes in a direct and immediate fashion. Like the famous 'trolley problem' thought experiment (Thomson, 1976), examining the logic and reasoning behind the ethical decision-making processes of AVs offers some of the most clear-cut cases of the ethical dilemmas we currently face.

This thesis will first explain the contemporary relevance of the technoethical dilemma of AVs and then demonstrate the interdisciplinary connectedness of philosophy (ethics) and the computer sciences concerning the issue (programming, designing, developing, engineering, etc..). The 's' at the end of 'computer sciences' is intended. It is almost a necessity to be well-versed in multiple disciplines, as, these days, a technologically

advanced end-result requires a combination of various expertise. For example, the designer needs the developer and vice versa.

This thesis will examine the ethical viewpoints relevant to the ethical-decision making processes of AVs from a philosophical perspective. This thesis will then demonstrate how an AV would react to an unavoidable crash scenario according to each ethical stance. Pseudocode for the ethical decision-making processes of each ethical setting will also be provided, in the programming language of Python, for further illustration and deeper understanding.

It is the goal of this thesis, that after reading, the reader will comprehend these ethical concepts and their significant contemporary relevance. These concepts need to be considered and understood by the general public and stakeholders in order for them to be able to participate in the greater conversation and for us to finally come to a general consensus on such matters.

2 Technoethics of Autonomous Vehicles

Technoethics is a field of study that examines the ethical implications and considerations arising from the development and implementation of technology. It involves evaluating the ethical impacts of technological advancements, exploring the moral dilemmas they present, and establishing guidelines for responsible and ethical practices. Technoethics aims to address the potential social, cultural, economic, and environmental consequences of technology, while promoting values such as privacy, security, transparency, accountability, fairness, and the well-being of individuals and society as a whole. It involves interdisciplinary research, incorporating elements of philosophy, ethics, sociology, law, policy, and technology studies to critically analyze and guide the ethical dimensions of technological innovation and application. (Luppicini, 2010)

The development of AVs has been a substantial factor in increasing the ethical urgency to answer age-old ethical conundrums. AVs are self-driving vehicles, driverless vehicles with incredible momentum, weighing tons and moving at fast speeds, interacting with humans on a daily, if not constant, basis. These autonomous entities have the power and the potential to inflict a devastating amount of mayhem and carnage.

One may ask, why have AVs at all if they are so potentially harmful? Beside the fact that technological progress is naturally evolving in this direction, about 90% of motor vehicle crashes can be reduced to human error, rather than other factors like vehicle malfunction (Litman, 2014). AVs can process information

magnatudes faster and more precisely than humans can. With sensors and processing power faster and far more accurate than our own, AVs can analyze the vehicle's wheel grip on the road and calculate risk of a slide and predict trajectory probabilities and make the exact appropriate corrections to the vehicle in milliseconds. A human, on the other hand, could be distracted and not even notice the oncoming danger. A human could make wrong assessments about the situation or make inappropriate corrections leading to an undesired outcome. According to some research, AVs could reduce crash rates 50% or even put an end to them altogether. (Garza, 2012)

AVs are obviously equipped with self-driving protocols to keep the occupants safe under normal circumstances. AVs are also equipped with ethical settings in the case of an unavoidable collision. In an unavoidable crash situation, where an AV must choose between colliding with a pedestrian or swerving and potentially causing harm to the vehicle's occupants, what should the AV be programmed to do? Should the AV always protect its occupants at all expense? Should the AV always try to reduce the total amount of damage or should the AV always try to minimize loss of life? What is the best ethical programming for AVs in an unavoidable crash situation in which human lives are at stake?

In the next sections, the focus will be on understanding the ethical reasoning and programming logic behind these ethical settings. This study will be conducted utilizing the interdisciplinary perspectives of philosophy (ethics) and computer science (programming).

3 Ethics of AV decision-making

In order to fully understand the ethical decision-making processes of AVs, one must understand the underlying ethical concepts and the theories behind them. The philosophical branch of ethics deals with the moral principles and ideals that govern one’s conduct. While ethics has yet to provide “one rule to rule them all”, ethics examines the concepts of right/good and wrong/bad, classifying and categorizing concepts, in a scientific and systematic way. Ethics can recommend different concepts of right and wrong based on one’s preferences and justifications. (Driver, 2022)

People come from all walks of life and have very different ethical viewpoints for various but equally valid reasons. Ethical views based on intuitive feelings or leap-of-faith assumptions would be a quite flimsy foundation from which to build an ethical framework that needs to be globally applicable. These ethical matters need to be reduced to their lowest common denominator to get a globally applicable result to be the

foundation for a universally applicable ethical framework. The only way to come to a global consensus is to base it on rational, logical, and empiric science (Petersen & Ryberg, 2010). This brings us to applied ethics and normative ethical theories.

Applied ethics and main normative ethical theories

Applied ethics deals with moral dilemmas in a pragmatic fashion. How can moral considerations be practically realized in the real world? Applied ethics deals in finding an answer to a moral dilemma we need an answer to for practical application. Applied ethics usually utilizes prima facie ethical principles and normative ethical theories to reach verdicts on ethically controversial issues that demand prompt real-world solutions. (Petersen & Ryberg, 2010)

Principlism, based on four prima facie ethical principles, is the most influential and widely utilized applied ethics approaches in bioethics and health care ethics. The prima facie principles of principlism are autonomy, non-maleficence, beneficence, and justice. (Beauchamp & Childress, 1994)

Normative ethical theories also play a significant role in attempting to form a universally applicable ethic. Normative ethical theories provide frameworks and principles that aim to guide ethical decision-making and evaluate the moral appropriateness of actions. When it comes to forming a universally applicable ethic, normative ethical theories offer different approaches that attempt to address the complexities of ethics in a broad sense (Garner & Rosen, 1967). By examining these theories, researchers and engineers can strive to design systems that make ethically informed decisions in ambiguous situations. The main normative ethical theories are deontology, virtue ethics, and consequentialism (Kagan, 1989). In Sections 3.1 through 3.3, these three main normative theories are briefly but concisely examined.

The ethical principles of ethical egoism and altruism are also examined in Sections 3.4 and 3.5, as it would be negligent to ignore the relevance of an individual's tendency for self-sacrifice, impartiality, or blatant self-interest when considering people's driving motives. Ethical egoism is basically self-interested, self-centered selfishness (Sanders, 1988). Altruism, on the other hand, is the ethical practice of concern for others and their welfare over one's own. The happiness of others will, according to altruism, result in an increase in the quality of life both physically and spiritually (Okasha, 2011). Egoism and altruism are included in this thesis as they are relevant ethical viewpoints, often mentioned in surrounding studies and literature. To base this claim, the variables and results of two studies on the subject (Contissa et al., 2017; Kumfer & Burgess, 2015) will be examined in Section 4.

After each ethical theory, there will be an explanation of how each of these ethical theories would translate into an AVs ethical setting's properties and priorities (i.e an AVs ethically based behavior on the road). We

will examine how an AV would make ethical decisions in an unavoidable crash scenario according to each ethical setting. At the end of each section, for illustrative purposes, these ethical principles will be translated into pseudocode, using the programming language of Python, followed by a concise explanation of the functions, instances, classes, and methods used in the code.

3.1 Deontology

A deontologist believes that one has a duty or an obligation to act in accordance with one's ethical maxim in every situation and that an action's motives have more moral value than an action's consequences. Deontology focuses on the inherent nature of an action and claims that certain actions are considered morally right or wrong regardless of their outcomes. Some of the tenets of deontology are a sense of duty, fairness, honesty, and respect for others. (Waller, 2005)

Immanuel Kant, the celebrated proponent of deontology, 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. Rationality is the main criterion in his formulation of *The Categorical Imperative* (initially published in 1785). According to Kant's Categorical Imperative, one should never treat someone as a means to an end but as ends in themselves. One should also only "act only according to that maxim whereby you can at the same time will that it should become a universal law" (Kant, 2003). Basically, the Categorical Imperative states that one should only act in the manner in which one would wish everyone to act all of the time.

3.1.1 Deontological ethical decision-making in an unavoidable crash situation

In the context of an unavoidable crash situation, a deontologist would base their decision on adherence to moral duties and principles rather than focusing on the consequences. A deontologist might approach the situation by identifying and applying a set of rules or principles that should guide their decision-making. These principles may include respect for human life, the duty to avoid causing harm, or adherence to traffic regulations and laws. The AV would be programmed to take actions that minimize harm to human beings, even if it means sacrificing the well-being of the occupants or other parties involved.

However, deontological principles can lead to rigid ethical decision-making and the potential for deadlock between conflicting moral duties. For instance, if the only options available are hitting a group of pedestrians

or swerving into oncoming traffic, a deontologist might face a moral dilemma where two principles, such as preserving human life and avoiding harm to others, come into conflict.

3.1.2 Deontology's decision-making code

The programming code for the deontological ethical setting in AVs can also vary depending on the specific implementation and programming language used. However, a simplified example in pseudocode to demonstrate the concept is provided in Code Segment 1. This example is for illustrative purposes and may not represent the actual implementation used in real-world AVs.

Code Segment 1. Deontological pseudocode using Python.

```
def make_decision(deontological_rules):
    # Determine the applicable rules for the situation
    applicable_rules = []
    for rule in deontological_rules:
        if rule.applies_to_current_situation():
            applicable_rules.append(rule)

    # Select the action that is allowed by the highest priority rule
    best_action = None
    highest_priority = -1
    for rule in applicable_rules:
        if rule.priority > highest_priority:
            best_action = rule.allowed_action
            highest_priority = rule.priority

    return best_action

# Define deontological rules
class DeontologicalRule:
    def __init__(self, applies_to_current_situation, allowed_action, priority):
        self.applies_to_current_situation = applies_to_current_situation
        self.allowed_action = allowed_action
        self.priority = priority

# Define specific deontological rules
def rule1_applies_to_current_situation():
    # Check if rule 1 applies to the current situation
    # Example: Always prioritize the safety of pedestrians over passengers
    return True

def rule1_allowed_action():
    # Define the allowed action for rule 1
    # Example: Always apply emergency braking to avoid hitting pedestrians
    return "Emergency Brake"

def rule1_priority():
    # Define the priority of rule 1
    # Example: Higher priority means it overrides lower priority rules
    return 1

def rule2_applies_to_current_situation():
    # Check if rule 2 applies to the current situation
    # Example: Always follow traffic rules and regulations
    return True
```

```
def rule2_allowed_action():
    # Define the allowed action for rule 2
    # Example: Follow the traffic light signal
    return "Follow Traffic Light"

def rule2_priority():
    # Define the priority of rule 2
    # Example: Lower priority compared to rule 1
    return 2

# Main code
deontological_rules = [
    DeontologicalRule(rule1_applies_to_current_situation, rule1_allowed_action, rule1_priority),
    DeontologicalRule(rule2_applies_to_current_situation, rule2_allowed_action, rule2_priority)
]
best_action = make_decision(deontological_rules)
execute_action(best_action)
```

Explanation of the code

In this example, the 'make_decision' function takes into account a set of deontological rules. It determines the applicable rules for the current situation by evaluating the 'applies_to_current_situation' method of each rule. Then, it selects the action allowed by the highest priority rule. The priority of each rule determines its precedence.

The deontological rules are defined as instances of the 'DeontologicalRule' class. Each rule consists of three components: the 'applies_to_current_situation' method, which checks if the rule is relevant to the current situation; the 'allowed_action' method, which specifies the action allowed by the rule; and the 'priority' method, which determines the priority of the rule.

The actual implementation of deontological ethical decision-making in AVs can be much more intricate, incorporating various factors and rules specific to the ethical framework being followed. The provided example offers a simplified representation to help demonstrate the concept of deontological ethical decision-making.

3.2 Virtue ethics

Virtue ethics, envisioned by Aristotle' in his seminal work, Nicomachean Ethics (first composed around 350 BC), is the ethical theory that emphasizes the development of moral character and the cultivation of virtues and elimination of vices. (Aristotle et al., 2011) Examples of virtues include empathy, responsibility, and prudence. As society is a collective of individuals, this individual cultivation of virtuous behavior will lead to

a flourishing of society as a whole, according to virtue ethics. 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)

3.2.1 Virtue ethic's ethical decision-making in an unavoidable crash situation

In the context of an unavoidable crash situation, a virtue ethicist would focus on the actions that align with virtuous characteristics and promote the overall flourishing of individuals and society. Instead of focusing on consequences (utilitarianism) or adhering to specific rules (deontology), a virtue ethicist would consider what a morally virtuous person would do in a given situation. Virtues such as compassion, empathy, responsibility, and fairness would guide the decision-making process.

A virtue ethicist would consider the virtues that are relevant to the situation and strive to act in accordance with those virtues. For example, they might prioritize minimizing harm, demonstrating empathy towards all affected parties, and taking responsibility for their actions, rather than following a predetermined set of rules or maximizing overall utility.

However, virtue ethics does not provide a specific set of instructions or guidelines for AVs in unavoidable crash situations. Implementing virtue ethics in AVs raises challenges in determining how to translate abstract virtues into concrete decision-making algorithms and addressing the potential conflicts that may arise when different virtues come into play.

3.2.2 Virtue ethics' decision-making code

The programming code for the virtue ethics setting in AVs can also vary depending on the specific implementation and programming language used. However, a simplified example in pseudocode to demonstrate the concept is in Code Segment 2. This example is for illustrative purposes and may not represent the actual implementation used in real-world AVs.

Code Segment 2. Virtue Ethics pseudocode using Python.

```
def make_decision(virtue_traits):
    # Determine the virtue-based evaluation for each potential action
    action_evaluations = []
    for action in possible_actions:
        evaluation = evaluate_action(action, virtue_traits)
        action_evaluations.append(evaluation)

    # Select the action with the highest virtue-based evaluation
    best_action = possible_actions[action_evaluations.index(max(action_evaluations))]

    return best_action
```

```
def evaluate_action(action, virtue_traits):
    # Evaluate the action based on virtue traits
    evaluation = 0

    # Consider virtue traits such as empathy, responsibility, and prudence
    evaluation += evaluate_empathy(action, virtue_traits)
    evaluation += evaluate_responsibility(action, virtue_traits)
    evaluation += evaluate_prudence(action, virtue_traits)

    return evaluation

def evaluate_empathy(action, virtue_traits):
    # Evaluate the action based on empathy virtue
    # Example: Consider the action's impact on the well-being and feelings of others
    evaluation = 0

    # Evaluate empathy virtue for the action

    return evaluation

def evaluate_responsibility(action, virtue_traits):
    # Evaluate the action based on responsibility virtue
    # Example: Consider the action's adherence to laws, regulations, and ethical standards
    evaluation = 0

    # Evaluate responsibility virtue for the action

    return evaluation

def evaluate_prudence(action, virtue_traits):
    # Evaluate the action based on prudence virtue
    # Example: Consider the action's ability to minimize risks and make sound judgments
    evaluation = 0

    # Evaluate prudence virtue for the action

    return evaluation

# Main code
virtue_traits = obtain_virtue_traits()
best_action = make_decision(virtue_traits)
execute_action(best_action)
```

Explanation of the code

In this example, the 'make_decision' function considers virtue traits and evaluates each potential action based on those traits. The 'evaluate_action' function assesses the action's virtue-based evaluation by considering specific virtue traits, such as empathy, responsibility, and prudence. The code assigns a numerical evaluation to each virtue trait for a given action and sums up those evaluations to determine the overall virtue-based evaluation.

The virtue traits and their associated evaluations are determined by the 'evaluate_empathy', 'evaluate_responsibility', and 'evaluate_prudence' functions. These functions evaluate the action's virtue-based considerations according to their respective virtue traits.

The actual implementation of virtue ethics in AVs can be more complex, involving additional virtue traits and intricate evaluations. The provided example aims to demonstrate the concept of virtue ethics in decision-making for AVs in a simplified manner.

3.3 Consequentialism - Utilitarianism

Consequentialism is the ethical theory that says an action's moral worth depends, entirely and solely, on its consequences. An action with an outcome that has more overall benefits than harm is good, while an action with an outcome that causes more harm than benefit is bad. The most famous and most applied version of consequentialism is utilitarianism. (Scheffler, 1988)

Utilitarianism is an ethical theory that suggests the best course of action is the one that maximizes universal happiness or has the greatest utility for the greatest number of people (Mill, 1863). Under utilitarianism, the end justifies the means. Individual rights, and even lives, can be ethically sacrificed and traded for the greater good. A utilitarian is expected to treat everyone equally and be impartial and judge ethical issues with a collective morality.

3.3.1 Utilitarian ethical decision-making in an unavoidable crash situation

In the context of an unavoidable crash situation, a utilitarian approach would prioritize minimizing harm and maximizing overall welfare. A utilitarian perspective would likely involve assessing the situation, predicting the potential consequences, and choosing the action that minimizes the overall loss of life and injury. This could mean making decisions based on factors such as the number of people involved, the severity of potential injuries, and the likelihood of different outcomes. A utilitarian AV might prioritize colliding with a single pedestrian to avoid hitting a group of pedestrians, based on the assumption that the potential harm caused to the group of pedestrians is greater than the harm caused to the individual.

However, it's important to note that implementing utilitarian principles in AVs is not without challenges and ethical dilemmas. Determining the exact course of action that maximizes overall welfare in complex real-world scenarios is quite difficult. How can one predict the entire reach of an action’s consequences when things are so intricately interconnected? How can, even the most powerful processor, ever predict how far the ripple of causality will travel through time and space with any degree of accuracy? How is one to predict the hurricane from the flapping of a butterfly’s wings?

3.3.2 Utilitarianism's decision-making code

The programming code for the utilitarian ethical setting in AVs can vary depending on the specific implementation and the programming language used. However, a simplified example in pseudocode to demonstrate the concept is in Code Segment 3. This example is for illustrative purposes and may not represent the actual implementation used in real-world AVs.

Code Segment 3. Utilitarian pseudocode using Python.

```
def make_decision(utilitarian_factors):
    # Determine the overall utility for each potential action
    action_utilities = []
    for action in possible_actions:
        utility = calculate_utility(action, utilitarian_factors)
        action_utilities.append(utility)

    # Select the action with the highest utility
    best_action = possible_actions[action_utilities.index(max(action_utilities))]

    return best_action

def calculate_utility(action, utilitarian_factors):
    # Calculate the utility for the given action based on utilitarian factors
    utility = 0

    # Consider factors such as safety, efficiency, and fairness
    utility += calculate_safety_utility(action)
    utility += calculate_efficiency_utility(action)
    utility += calculate_fairness_utility(action, utilitarian_factors)

    return utility

def calculate_safety_utility(action):
    # Calculate the utility based on safety considerations
    # Example: Consider factors like minimizing risk of harm to passengers, pedestrians, and other vehicles
    utility = 0

    # Calculate safety utility based on action

    return utility

def calculate_efficiency_utility(action):
    # Calculate the utility based on efficiency considerations
    # Example: Consider factors like minimizing travel time, reducing traffic congestion
    utility = 0

    # Calculate efficiency utility based on action

    return utility

def calculate_fairness_utility(action, utilitarian_factors):
    # Calculate the utility based on fairness considerations
    # Example: Consider factors like equal distribution of benefits and minimizing harm to disadvantaged groups
    utility = 0

    # Calculate fairness utility based on action and utilitarian factors

    return utility

# Main code
```

```
utilitarian_factors = obtain_utilitarian_factors()
best_action = make_decision(utilitarian_factors)
execute_action(best_action)
```

Explanation of the code

In this example, the 'make_decision' function takes into account utilitarian factors, such as safety, efficiency, and fairness. It calculates the utility for each potential action using the 'calculate_utility' function, which considers these factors and their associated calculations. Finally, the code selects the action with the highest utility and executes it.

The actual implementations of ethical decision-making in AVs can be much more complex, involving extensive sensor data processing, machine learning algorithms, and integration with traffic regulations. The example provided here is a simplified representation to help illustrate the concept of utilitarian ethical decision-making.

3.4 Ethical egoism

Ethical egoism is the ethical theory that asserts individuals should act in their own self-interest. According to ethical egoism, individuals ought to embrace selfishness, and make decisions that maximize their own well-being, regardless of the consequences and impact on others (Shaver, 2019). In other words, the best thing for you, is the right thing to do. Ethical egoism, as selfishness, is a vice according to virtue ethics. Selfishness contrasts the ethical mindset of the utilitarian ethic. Selfishness is also unethical from a deontological perspective as well, because it fails to treat others as ends in themselves and contradicts the idea of universal moral principles. (Shaver, 2019)

This perspective may seem unethical, but doing what's best for oneself, being selfish to a degree, is a necessary part of survival. Survival of self and kin is an ingrained primal instinct, not only of humans and animals but of all life. From this perspective, egoistic selfish survival could be viewed as the only rule. Egoism, therefore, can condone an "eat or be eaten", "survival of the fittest or most ruthless" mindset. (Sanders, 1988)

3.4.1 Ethical egoism's ethical decision-making in an unavoidable crash situation

In the context of an unavoidable crash situation, an ethical egoist would prioritize the well-being and self-preservation of the vehicle's occupants. Under ethical egoism, the AV would be programmed to prioritize protecting the lives and minimizing harm to its occupants above all else. In an unavoidable crash situation, the AV would prioritize actions that maximize the chances of survival for the individuals inside, even if it means potentially causing harm to pedestrians, other drivers, or passengers in other vehicles.

However, ethical egoism does not take into account the well-being or interests of others, and the primary concern is the self-interest of the vehicle's occupants. This approach does not consider broader moral considerations or the well-being of society as a whole. Prioritizing the lives of the AVs occupants at the expense of others can lead to conflicts with other ethical principles, societal expectations, and legal frameworks.

3.4.2 Ethical egoism's decision-making code

In the context of ethical decision-making processes for AVs, the ethical egoist perspective focuses on maximizing the self-interest of the vehicle and its occupants. A simplified example to demonstrate the concept in pseudocode is in Code Segment 4. This example is for illustrative purposes and may not represent the actual implementation used in real-world AVs.

Code Segment 4. Ethical Egoism pseudocode using Python.

```
def make_decision(egoist_factors):
    # Determine the egoistic evaluation for each potential action
    action_evaluations = []
    for action in possible_actions:
        evaluation = evaluate_action(action, egoist_factors)
        action_evaluations.append(evaluation)

    # Select the action with the highest egoistic evaluation
    best_action = possible_actions[action_evaluations.index(max(action_evaluations))]

    return best_action

def evaluate_action(action, egoist_factors):
    # Evaluate the action based on egoistic factors
    evaluation = 0

    # Consider egoistic factors such as self-preservation and maximizing benefits to the vehicle's occupants
    evaluation += evaluate_self_preservation(action, egoist_factors)
    evaluation += evaluate_occupant_benefits(action, egoist_factors)

    return evaluation

def evaluate_self_preservation(action, egoist_factors):
    # Evaluate the action based on self-preservation
    # Example: Consider the action's ability to avoid harm or damage to the vehicle
    evaluation = 0

    # Evaluate self-preservation for the action
```

```
    return evaluation

def evaluate_occupant_benefits(action, egoist_factors):
    # Evaluate the action based on maximizing benefits to the vehicle's occupants
    # Example: Consider the action's ability to prioritize comfort and convenience for the occupants
    evaluation = 0

    # Evaluate occupant benefits for the action

    return evaluation

# Main code
egoist_factors = obtain_egoist_factors()
best_action = make_decision(egoist_factors)
execute_action(best_action)
```

Explanation of the code

In this example, the 'make_decision' function considers egoistic factors and evaluates each potential action based on those factors. The 'evaluate_action' function assesses the action's egoistic evaluation by considering specific egoistic factors, such as self-preservation and maximizing benefits to the vehicle's occupants. The code assigns a numerical evaluation to each egoistic factor for a given action and sums up those evaluations to determine the overall egoistic evaluation.

The egoist factors and their associated evaluations are determined by the 'evaluate_self_preservation' and 'evaluate_occupant_benefits' functions. These functions evaluate the action's egoistic considerations according to their respective factors.

The provided example demonstrates a simplified representation of the ethical egoist setting for decision-making in AVs. The ethical egoist perspective in AVs is focused on prioritizing self-interest, which may conflict with other ethical frameworks that prioritize the greater good or other moral considerations.

3.5 Altruism

Altruism is an ethical viewpoint that emphasizes acting for the benefit and well-being of others, even at the expense of one's own interests. According to altruism, giving of oneself and caring for others is always the ethical thing to do, no matter what one's own needs are in any given situation (Zahavi, 1995). Altruism is the opposite end of the ethical spectrum of ethical egoism/selfishness. Where an egoist may risk others' health

for their own advantage, an altruist may give up too much of themselves and neglect their own health or social needs in order to care for others.

Some individuals could potentially prefer an altruistic ethical setting in their AV. For example, single passenger AVs containing a passenger that is: impartial, terminally ill, depressed, guilt ridden, grief stricken, suicidal, give up the will to live, transporting a criminal or patient on life-support, etc… This altruistic setting would, theoretically, not interfere or conflict with any of the other AV ethical settings real world implications. (Rosebury, 2021)

Altruism, while emphasizing the well-being of others, does not completely ignore the self-interest of the vehicle's occupants. In an altruistic approach, the well-being of the occupants is considered within the broader context of maximizing overall welfare and minimizing harm, much like the tenets of utilitarianism. (MacAskill, 2017)

3.5.1 Altruistic ethical decision-making in an unavoidable crash situation

In the context of an unavoidable crash scenario, an altruistic approach would prioritize the well-being and safety of all parties involved, seeking to minimize harm and maximize overall welfare. According to altruism, the AV would be programmed to make decisions that prioritize the preservation of human life and the well-being of all individuals, including both the occupants of the vehicle and any pedestrians or occupants of other vehicles.

Much like the utilitarian approach, an altruistic AV would aim to take actions that minimize harm to all parties involved, potentially considering factors such as the number of individuals affected, the severity of potential injuries, and the likelihood of different outcomes. The goal would be to make decisions that maximize overall welfare, without favoring any particular individual or group.

Implementing altruism in AVs can be challenging as it requires making difficult trade-offs and balancing conflicting interests. For instance, in situations where multiple collision scenarios are possible, determining the course of action that minimizes harm to all parties can be complex and context-dependent.

3.5.2 Altruisism’s decision-making code

In the context of ethical decision-making processes for AVs, the altruist perspective focuses on prioritizing the well-being and safety of others over the vehicle and its occupants. A simplified example of what the programming code for the altruist ethical setting might look like is provided in Code Segment 5.

Code Segment 5. Altruist pseudocode using Python.

```
def make_decision(altruist_factors):
    # Determine the altruistic evaluation for each potential action
    action_evaluations = []
    for action in possible_actions:
        evaluation = evaluate_action(action, altruist_factors)
        action_evaluations.append(evaluation)

    # Select the action with the highest altruistic evaluation
    best_action = possible_actions[action_evaluations.index(max(action_evaluations))]

    return best_action

def evaluate_action(action, altruist_factors):
    # Evaluate the action based on altruistic factors
    evaluation = 0

    # Consider altruistic factors such as prioritizing the safety and well-being of others
    evaluation += evaluate_safety_of_others(action, altruist_factors)
    evaluation += evaluate_minimizing_harm(action, altruist_factors)

    return evaluation

def evaluate_safety_of_others(action, altruist_factors):
    # Evaluate the action based on the safety and well-being of others
    # Example: Consider the action's ability to prioritize pedestrians' and other drivers' safety
    evaluation = 0

    # Evaluate safety of others for the action

    return evaluation

def evaluate_minimizing_harm(action, altruist_factors):
    # Evaluate the action based on minimizing harm to others
    # Example: Consider the action's ability to minimize potential harm or damage to others
    evaluation = 0

    # Evaluate minimizing harm for the action

    return evaluation

# Main code
altruist_factors = obtain_altruist_factors()
best_action = make_decision(altruist_factors)
execute_action(best_action)
```

Explanation of the code

In this example, the make_decision function considers altruistic factors and evaluates each potential action based on those factors. The evaluate_action function assesses the action's altruistic evaluation by considering specific altruistic factors, such as the safety and well-being of others and minimizing harm to others. The code assigns a numerical evaluation to each altruistic factor for a given action and sums up those evaluations to determine the overall altruistic evaluation.

The altruist factors and their associated evaluations are determined by the evaluate_safety_of_others and evaluate_minimizing_harm functions. These functions evaluate the action's altruistic considerations according to their respective factors.

The altruist perspective in AVs prioritizes the well-being and safety of others, which may involve sacrificing the vehicle's occupants' interests. The provided example demonstrates a simplified representation of the altruist setting for decision-making in AVs.

4 Relevant studies on AV ethical decision-making

The two studies (Contissa et al., 2017; Kumfer & Burgess, 2015) mentioned earlier were included to elucidate and compare these ethical concepts and show how they have been relevant and significant to the ongoing greater conversation. It is not the intention of this thesis to fully analyze the parameters, methods, or results of these studies. These studies are briefly mentioned for clarification and so that the reader will better understand the practical implementation and implications of these ethical concepts.

4.1 Investigation into the role of rational ethics in crashes of automated vehicles

A recent study, *Investigation into the role of rational ethics in crashes of automated vehicles*, Kumfer & Burgess created an ethical thought experiment using MATLAB computer simulations (Kumfer & Burgess, 2015). They were able to calculate the average fatalities per ethics system per simulation. These simulations provided statistics of the virtual consequences of each of these ethical settings in an unavoidable crash scenario.

Included in this MATLAB study were the ethical settings of ethical egoism, utilitarianism, RFP, and virtue ethics. RFP ethics is basically a deontological ethical viewpoint. In RFP ethics, people are to be respected because of their intrinsic value, just as Kant's Categorical Imperative stipulates, one is to treat others as 'ends in themselves'. (Kumfer & Burgess, 2015) After 100,000 simulations, their study provides comparative statistics of the causal effect an AVs ethical settings could have in the real world. Certain basic parameters were assumed for the study, based on literature, such as age related to probability of fatality in a collision.

TABLE 1 Average Fatalities per Ethics System per Simulation

Simulation Variable	Number, by Simulation Number										Average
	1	2	3	4	5	6	7	8	9	10	
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

As can be seen in Table 1, the average fatalities per ethics system per simulation was calculated after ten rounds of simulations. This study used four out of five of the ethical viewpoints closely examined in this thesis. The average on the far right in the red box shows the average amount of fatalities after ten rounds of simulations. The utilitarian ethical setting, highlighted by the green rectangle, produced the least fatalities. The virtue ethics setting, on the other hand, produced the most fatalities. These results suggest that a utilitarian ethical setting, more than any other, may provide the greatest overall benefits to society. (Kumfer & Burgess, 2015)

4.2 The Ethical Knob: ethically-customisable automated vehicles and the law

The study, *The Ethical Knob: ethically-customisable automated vehicles and the law*, explored the possibility of an AV user having the freedom, or burden, of deciding their own personal ethics setting (PES) with an "Ethical Knob" (Contissa et al., 2017). As can be seen in Fig 1, this interactive knob gives the AV user the option to select one of three settings: altruistic (preference for third parties), impartial (equal importance given to occupants & third parties), and egoistic (preference for occupants).

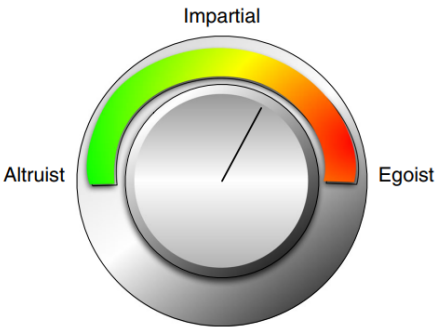

Fig. 1 The Ethical Knob (Contissa et al., 2017)

As can be seen from Fig. 1, only the ethical settings of altruist, egoist, or impartial were included in this study as considerable alternatives for an AV's ethical settings. Based on these two relevant studies on AV ethical decision-making, one can understand why the ethical views of altruism, egoism, and the other

normative ethical theories of this thesis are considered significant and should be included in examinations of similar subject matter.

5 Discussion

The intention of this thesis is to examine, compare, and understand the different ethical decision-making processes of AVs, not to judge the validity of these ethical viewpoints or to propose that an ethical viewpoint is more appropriate than another.

In this thesis, the main ethical theories relevant to the ethical decision-making of AVs (deontology, virtue ethics, utilitarianism, egoism, and altruism) were examined. Corresponding pseudocode for each ethical theory was given to illustrate theoretical implementation of these theories in the ethical decision-making of AVs. The manner in which each pseudocode segment corresponds to its representative ethical viewpoint will be discussed next.

In Code Segment 1, the deontology function takes into account sets of rules (i.e. duties and obligations). The function determines the applicable rules for a certain situation, gives priority to these rules, and then executes the action allowed by highest priority. By adhering to rules, duties, obligations, and virtuous behavior, this pseudocode corresponds to the tenets of deontology. (Waller, 2005)

In Code Segment 2, the virtue ethics function assesses and considers specific virtue traits, such as empathy, responsibility, and prudence. The code attempts to evaluate the maximum sum of all possible virtuous actions and adhere to what produces the most virtuous outcome. By prioritizing virtues and adhering to them in an unavoidable crash scenario, the virtue ethics pseudocode corresponds to the tenets of virtue ethics (Hursthouse & Pettigrove, 2022). Unfortunately, this ethical setting had the most fatalities in the MATLAB computer simulations. (Kumfer & Burgess, 2015)

In Code Segment 3, the utilitarian function takes into account factors, such as safety, efficiency, and fairness and calculates the potential utility for each potential action. The code selects the action with the highest utility and executes it. The utilitarian ethical setting had the least fatalities in the MATLAB computer simulations (Kumfer & Burgess, 2015). By prioritizing and maximizing overall utility, this pseudocode adheres and corresponds to the tenets of utilitarianism. (Mill, 1863)

It's important to note that these normative ethical theories have been subject to criticism and debate. Critics argue that some theories may neglect the importance of pluralistic and relativistic factors such as cultural, contextual, and subjective factors in ethical decision-making of AVs. Additionally, different theories may offer conflicting principles or guidelines, leading to disagreements about what constitutes a universally applicable ethic. (Bagg, 2016; Scheffler, 1988)

In Code Segment 4, the egoist function considers factors, such as self-preservation and maximizing benefits to the vehicle's occupants. The code ranks each egoistic factor for a given action and sums up those evaluations to determine the overall most beneficial action for the egoist. By being self-interested and maximizing personal gain, the pseudocode adheres and corresponds to the tenets of ethical egoism. (Shaver, 2019)

Egoism and utilitarianism can seem like incompatible ethical viewpoints. However, it may be the case that what most benefits the egoist, also indirectly coincides with what produces the most overall utility. In this way, while contrasting utilitarianism, ethical egoism may not necessarily contradict or completely undermine the utilitarian ethic. (Sanders, 1988)

In Code Segment 5, the altruist function prioritizes safety and well-being of others and minimizes harm to others at all costs. The code ranks altruistic factors for any given action and sums up those evaluations to determine and execute the action with the most overall altruistic value. In this way, the pseudocode adheres and corresponds to the tenets of altruism (Zahavi, 1995). While the ethical views of utilitarianism and altruism share similar tenets, they differ in their scope and focus. Utilitarianism may prioritize actions that produce the most net benefits, even if they require sacrificing the well-being of a few individuals for the greater good, while altruism tends to prioritize immediate acts of compassion and empathy on an individual level. (Rosebury, 2021)

While the ethical theories of this thesis provide valuable insights, it's important to acknowledge that applying them to the complex real-world context of AVs will always involve challenges and trade-offs. Initially, practical implementation will no doubt be a turbulent process and may require balancing conflicting ethical principles, addressing cultural differences, and considering public opinion (Litman, 2014). Hopefully, future programmers, designers, philosophers, policy makers, stakeholders, and the like, will consider these sentiments and take them to heart.

6 Conclusion

Due to technological evolution and new advancements in autonomous technology, it is more important than ever that we understand the world around us and the concepts that, literally, drive us forward. In this thesis, we have interdisciplinarily demonstrated the ethical reasoning and the programming logic behind the different ethical decision-making processes of AVs from a philosophical perspective and a computer programming perspective.

In this thesis. five ethical theories and decision-making processes were examined. The logic and ethical reasoning for each ethical viewpoint were contrasted and analyzed. This logic and ethical reasoning was then further illustrated in the programming language of Python. This thesis attempted to demonstrate how to implement these philosophical theories from natural language into logical programming language.

Each ethical theory, each valid in its own way, has a different approach and a different intended end result in the case of an unavoidable crash scenario. It is important to understand these concepts, and the reasoning behind them, as these ethical concepts are crucial and central to understanding and participating in the greater debate currently surrounding the realm of autonomous processes and autonomous entities, not only AVs.

By considering these theories, the stakeholders, like the general public, developers, and policy makers of AVs can address public concerns and expectations regarding ethical decision-making and come to a general consensus. Engaging with established ethical frameworks can help increase transparency, accountability, and public trust in autonomous technologies.

Finally, interdisciplinary collaboration involving ethicists, technologists, policymakers, and the public is crucial for ensuring socially responsible ethical decision-making processes in the development of AVs. Achieving a consensus and a truly universally applicable ethic may require ongoing dialogue, critical reflection, and a nuanced understanding of the complexities involved in the ethical decision-making of autonomous processes.

References

Aristotle, Bartlett, R. C., & Collins, S. D. (2011). Aristotle's Nicomachean ethics. University of Chicago Press.

Bagg, S. (2016). Between Critical and Normative Theory: Predictive Political Theory as a Deweyan Realism. Political Research Quarterly, 69(2), 233–244. https://doi.org/10.1177/1065912916634898

Beauchamp, T. L. and Childress, J. F. (1994) Principles of medical ethics. New York: Oxford University Press.

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

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

Driver, J. (2002). "Moral Theory", The Stanford Encyclopedia of Philosophy (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), Visited 01 May 2023.

Garner, R. T., Rosen, B. (1967). Moral Philosophy: A Systematic Introduction to Normative Ethics and Meta-ethics. New York: Macmillan. p. 70.

Garza, A. P. (2012) “Look Ma, No Hands!”: Wrinkles and Wrecks in the Age of Autonomous Vehicles. New England Law Review, Vol. 46, 2012, pp. 581–616.

Gogoll, J., & Müller, J. F. (2017). Autonomous Cars: In Favor of a Mandatory Ethics Setting. Science and engineering ethics, 23(3), 681–700. https://doi.org/10.1007/s11948-016-9806-x

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

Kagan, S. (1989). The Limits of Morality. Oxford: Clarendon Press. p. 13.

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

Kumfer, W., & Burgess, R. (2015). Investigation into the Role of Rational Ethics in Crashes of Automated Vehicles. Transportation Research Record, 2489(1), 130–136. https://doi.org/10.3141/2489-15

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

Luppicini, R. (2010). Technoethics and the Evolving Knowledge Society: Ethical Issues in Technological Design, Research, Development, and Innovation. Advances in Information Security, Privacy, and Ethics. IGI Global. doi:10.4018/978-1-60566-952-6.

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/.

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

Petersen, T. S., & Ryberg, J. (2010). Applied Ethics. obo in Philosophy. doi: 10.1093/obo/9780195396577-0006

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.

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: Concepts, Viewpoints, Cases, and Codes. National Institute for Engineering Ethics, Lubbock, Tex.

Thomson J. J. (1976). Killing, letting die, and the trolley problem. The Monist, 59(2), 204–217. https://doi.org/10.5840/monist197659224

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

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