Technische Universität München

TUM School of Engineering and Design

An Ethical and Risk-aware Framework for Motion Planning of Autonomous Vehicles

Maximilian M. H. Geisslinger

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:	Prof. Dr. Constantinos Antoniou
Prüfende der Dissertation:	1. Prof. Dr.-Ing. Markus Lienkamp
	2. Prof. Dr.-Ing. Christoph Stiller
	3. Prof. Derek Leben, Ph.D.

Die Dissertation wurde am 25.10.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 10.04.2024 angenommen.

Meinen Eltern

Acknowledgments

This dissertation was composed as a research associate at the Institute of Automotive Technology at the Technical University of Munich from 2019 to 2023, working on the ANDRE (Autonomous Driving Ethics) project in collaboration with the Institute for Ethics in Artificial Intelligence (IEAI).

First and foremost, I express my deepest gratitude to my professor and chair, Prof. Dr. Markus Lienkamp, for his expertise, exceptional guidance, and great support. Thank you very much for the liberty and trust you placed in me, fostering a climate of autonomy and creativity that enabled this work and essentially contributed to my professional and personal development.

Furthermore, I extend my sincere appreciation to my additional examiners, Prof. Dr. Christoph Stiller and Prof. Derek Leben, Ph.D., as well as Prof. Dr. Constantinos Antoniou, for taking the chairmanship of the examination.

Moreover, my gratitude belongs to my proofreaders, Dr. Alexander Wischnewski, Phillip Karle, Rainer Trauth, and Ilir Tahiraj, whose eyes for detail and commitment to ensuring clarity greatly enhanced the quality of the work. I also thank Sebastian Huber for his invaluable support with LaTeX.

To my esteemed project partner, the IEAI, and in particular, Franziska Poszler, I am indebted for your trustful collaboration that transcended disciplinary boundaries, contributing significantly to the success of my dissertation.

I am profoundly grateful for the collaborative experience within my research group, Intelligent Vehicle Systems and later Autonomous Vehicle Perception, and the exceptional individuals with whom I had the privilege of working in pursuing the pinnacle of autonomous driving. Your relentless passion and dedication are a true source of inspiration. Thanks to you, work hardly felt like work most of the time. I am especially proud to call many of you my friends beyond our work at the institute. In particular, I would like to thank my girlfriend, Maria Wolf, for her persistent support and warmth, even during the colder days.

Finally, I extend my deepest thanks to my family, with special appreciation to my parents. Your unwavering support has been instrumental, and I am forever thankful for your continual help. I dedicate this work to you.

Garching, October 2023

Maximilian Geisslinger

Contents

List of Abbreviations

ADS	Autonomous Driving System
AI	Artificial Intelligence
AV	Autonomous Vehicle
BC	Behavior Cloning
BEV	Birds-Eye-View
CAV	Connected Automated Vehicle
CI	Crash Index
CNN	Convolutional Neural Network
COVID	Corona Virus Disease
CPU	Central Processing Unit
CTRA	Constant Turn Rate Acceleration
CV	Constant Velocity
EU	European Union
GAIL	Generative Adversarial Imitational Learning
IRL	Inverse Reinforcement Learning
LQG	Linear-Quadratic Gaussian
LSTM	Long-Short-Term-Memory
MAIS	Maximum Abbreviated Injury Scale
ML	Machine Learning
MPC	Model Predictive Control
NHTSA	National Highway Traffic Safety Administration's Crash Report Sampling System
NLL	Negative Log Likelihood
ODD	Operational Design Domain
OEM	Original Equipment Manufacturer
PCC	Pearson Correlation Coefficient
PDF	Probability Density Function
POMDP	Partially Observable Markov Decision Process
RGB	Red-Green-Blue
RL	Reinforcement Learning
RNN	Recurrent Neural Network
RRT	Rapidly Exploring Random Tree
RSS	Responsibility Safety Shield
TET	Time Exposed Time-to-Collision
TTC	Time-to-Collision
TTR	Time-to-React
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
VRU	Vulnerable Road User

Formula Symbols

Formula Symbols	Unit	Description
a	m/s^2	Total acceleration
a_x	m/s^2	Acceleration in x
a_y	m/s^2	Acceleration in y
c	$-$	Empiric coefficient in harm model
d	m	Lateral displacement from a reference path
H	$-$	Estimated personal harm due to a collision
I_q	$-$	Inequality describing a question in the Ethical Vehicle Experiment
J_{total}	$-$	Total costs
J_{A}	$-$	Acceleration costs
J_{J}	$-$	Jerk costs
J_{SA}	$-$	Steering angle costs
J_{SR}	$-$	Steering rate costs
J_{EN}	$-$	Energy costs
J_{Y}	$-$	Yaw rate costs
J_{LC}	$-$	Lane center offset costs
J_{V}	$-$	Velocity offset costs
J_{O}	$-$	Orientation offset costs
J_{D}	$-$	Distance to obstacles costs
J_{L}	$-$	Path length costs
J_{ID}	$-$	Inverse duration costs
J_{Risk}	$-$	Risk costs
J_{Mobility}	$-$	Mobility costs
J_{Comfort}	$-$	Comfort costs
J_{B}	$-$	Bayes costs
J_{E}	$-$	Equality costs

J_{M}	$-$	Maximin costs
J_{R}	$-$	Responsibility costs
J_{S}	$-$	Selfish costs
m	kg	Mass
NLL	$-$	Negative Log Likelihood
$p_{\text{Collision}}$	$-$	Collision probability
$P_{\text{MAIS3+}}$	$-$	Probability for severity MAIS3+
PCC	$-$	Pearson Correlation Coefficient
q	$-$	Question in the Ethical Vehicle Experiment
R	$-$	Risk
R_{max}	$-$	Maximum acceptable risk
r	$-$	Parameter smaller than 1 describing responsibility
s	m	Arc length along a reference path
S_H	$-$	Set of harms
S_R	$-$	Set of risks
S_S	$-$	Set of scenarios
t	s	Time
t_0	s	Initial time step
t_f	s	Final time step
u	$-$	Trajectory
U	$-$	Set of trajectories
v	m/s	Velocity
w	$-$	Weighting parameter, indices analog to J
W	$-$	Solution space of weighting parameters
x_{GT}	m	Ground truth value in x
y_{GT}	m	Ground truth value in y
α	rad	Collision angle
γ	$-$	Empiric time discount factor
δ	rad	Steering angle
ΔJ	$-$	Difference in costs
μ_x	m	Expected position value in x
μ_y	m	Expected position value in y
Ψ	rad	Yaw angle

ρ	$-$	Correlation coefficient
σ_x	m	Standard deviation in x
σ_y	m	Standard deviation in y
Σ	$-$	Covariant matrix
$\mathcal{G}_S$	$-$	Goal region
$\mathcal{L}$	$-$	Loss
$\mathcal{W}_{S,\text{free}}$	$-$	Free space

1 Introduction

1.1 Motivation

Autonomous Vehicles (AVs) are being developed to increase the safety of roadways and transportation systems. They are expected to profoundly revolutionize our travel and transportation patterns [1, 2]. However, as recent events of fatal accidents with AVs involved [3] have indicated, it becomes apparent that even with their advanced capabilities, AVs may still be involved in accidents. Some of these accidents could have fatal consequences. There is an inherent risk associated with road traffic that cannot be completely mitigated by AVs [4]. This increases the challenge of AVs from a purely technical task by an ethical component: Ultimately, AVs will have to make decisions that were formerly the preserve of humans. Some of these decisions are expected to raise moral questions [5, 6].

Unlike human drivers who rely on instinct [7], AVs must be pre-programmed to handle critical situations that may involve life-or-death decisions. A prominent example of such a decision, which is often part of the public discourse, is the trolley problem [8]: In a hypothetical scenario, a person must make a life-or-death decision in controlling a runaway trolley. The dilemma arises when the person faces the choice of either allowing the trolley to continue on its current path, leading to multiple (e.g., five) fatalities, or diverting it onto another track, causing the death of one individual to save the others. Figure 1.1 gives a visualization of this widely discussed dilemma.

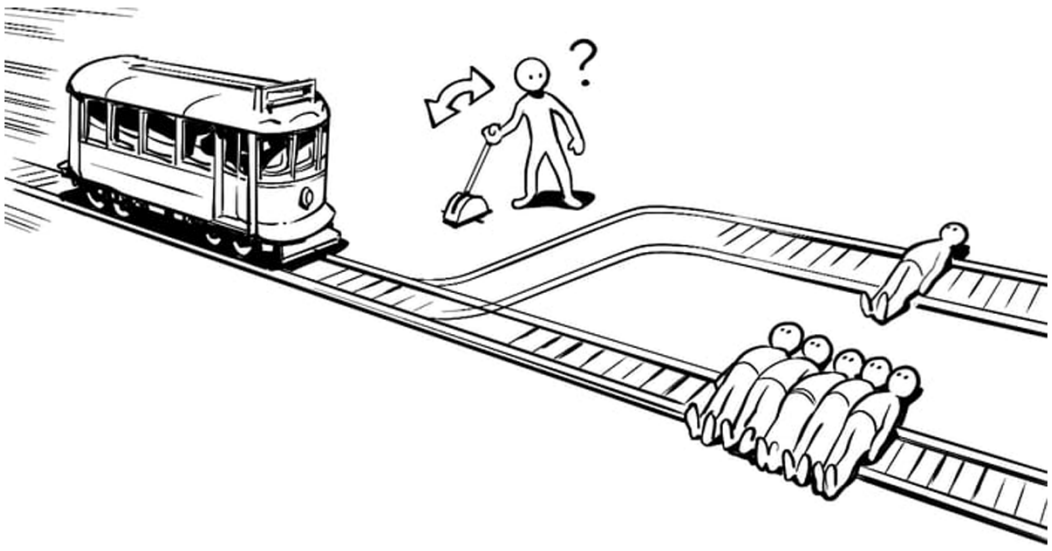


Figure 1.1: Visualization of the trolley problem, where a decision must be made that either lets a single person die or kills five persons. Retrieved from [9].

The core of this ethical dilemma is often transferred to AVs, which means that AVs must be equipped to face scenarios in which the safety of at least one person is at stake. This raises complex questions about prioritizing lives, minimizing harm, and adhering to social values. Determining how AVs should respond in such situations becomes a significant challenge, as it requires striking a delicate balance between utility, fairness, and individual safety. Ethics in autonomous driving has been shown to play a significant role in affecting public acceptance of this emerging technology [10]. The way AVs make decisions in complex situations can have a profound impact on how people perceive and trust these vehicles.

Therefore, expert groups [11], ethical committees [12], as well as legislation [13], have formulated rules and guidelines to address the ethical challenges associated with AVs. Many of these recommendations and guidelines affect the way in which the AV software must be programmed, especially when AVs make decisions autonomously. Developing algorithms that consider both technical and ethical aspects is still a significant challenge, where industry and research have not yet provided solutions. To the best of the author's knowledge, no algorithm for AV decision-making considers ethical aspects and is based on these guidelines. Developing such an algorithm that reflects the ethical challenges in AV decision-making is, thus, a crucial part of advancing the effort to bring AVs to the public streets.

1.2 Purpose of the Research

The objective of this thesis is to develop a trajectory planning algorithm for AVs that incorporates ethical considerations, aligning with recent legislation and guidelines. A key principle of all these guidelines is to avoid dilemma situations in AVs and instead manage risk using ethical principles. This shifts the focus from making life-and-death decisions, akin to trolley problems, to addressing questions of risk distribution.

The challenge of risk distribution is depicted in Figure 1.2 in a simplified way: The AV in blue is driving in the middle between a truck on the left and a cyclist on the right. The lateral positioning of the AV is assumed to affect the risks of road users. Moving closer to the cyclist increases the risk for the cyclist due to potentially severe injuries in case of an accident. Conversely, reducing the distance to the truck increases the risk for the AV, considering the higher consequences of a collision due to the different masses involved.

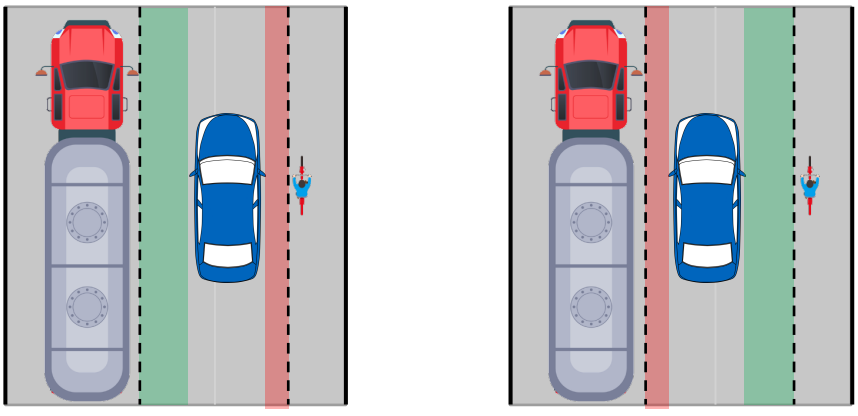


Figure 1.2: Simplified scheme to showcase risk distribution as a result of the AV trajectory. Depending on the lateral position of the blue av, the risks shift between the AV and the cyclist. The figure was modified according to [14].

This raises the fundamental question of how the AV should select its lateral position or, more generally, its trajectory. Research indicates that although people agree that the AV should minimize the overall risk, they tend to prefer AVs that prioritize their own safety [15]. This could be an incentive for Original Equipment Manufacturers (OEMs) and AV providers to program their vehicles with priority for the safety of passengers, paying less attention to other road users. As a result, the number of accidents between trucks and AVs could decrease faster than other types of accidents [16]. Figure 1.2 suggests that this self-centered behavior could have negative implications for Vulnerable Road Users (VRUs). This conflicts with the requirements of the European Union (EU) Commission and current legislation.

Considering that AVs should not selfishly distribute risks, this thesis explores how a fair distribution of risks could be achieved and effectively represented in an algorithm for trajectory planning. This constitutes the primary focus of the subsequent research.

1.3 Structure of the Thesis

This thesis presents an interdisciplinary approach to address the problem of ethical decision-making. Figure 1.3 shows a scheme of the structure of the thesis. In the following Chapter 2, the related works from the state of the art are presented. Due to the interdisciplinary core of this work, the state of the art is composed of two major disciplines, namely ethics and motion planning. The field of ethics covers relevant work from normative and descriptive ethics, which form the basis for the ethical considerations in the Methods sections. As the second discipline, the state of the art in motion planning is separated into the trajectory prediction part and a trajectory planning part. Chapters 3 and 4 present the underlying methods for developing an ethical algorithm. Chapter 3 presents an ethical framework for a trajectory planning algorithm, which combines the ideas of motion planning and ethics. Therefore, in the sense of normative ethics, a distribution of risks according to ethical principles is formulated and integrated into an algorithm. Chapter 4 presents a user experiment that investigates the participants' preferences of ethical principles within the framework, which relates to descriptive ethics. The results of the experiment provide weighting parameters to accomplish the algorithm. The parameterized algorithm is investigated in Chapter 5. Therefore, the algorithm is evaluated in a simulation of 2,000 scenarios with a focus on the empiric effects compared to the state of the art. Chapter 6 discusses the proposed framework in light of the posed requirements and research questions. Finally, this chapter gives some recommendations on how to proceed in establishing ethical considerations in AV software development.

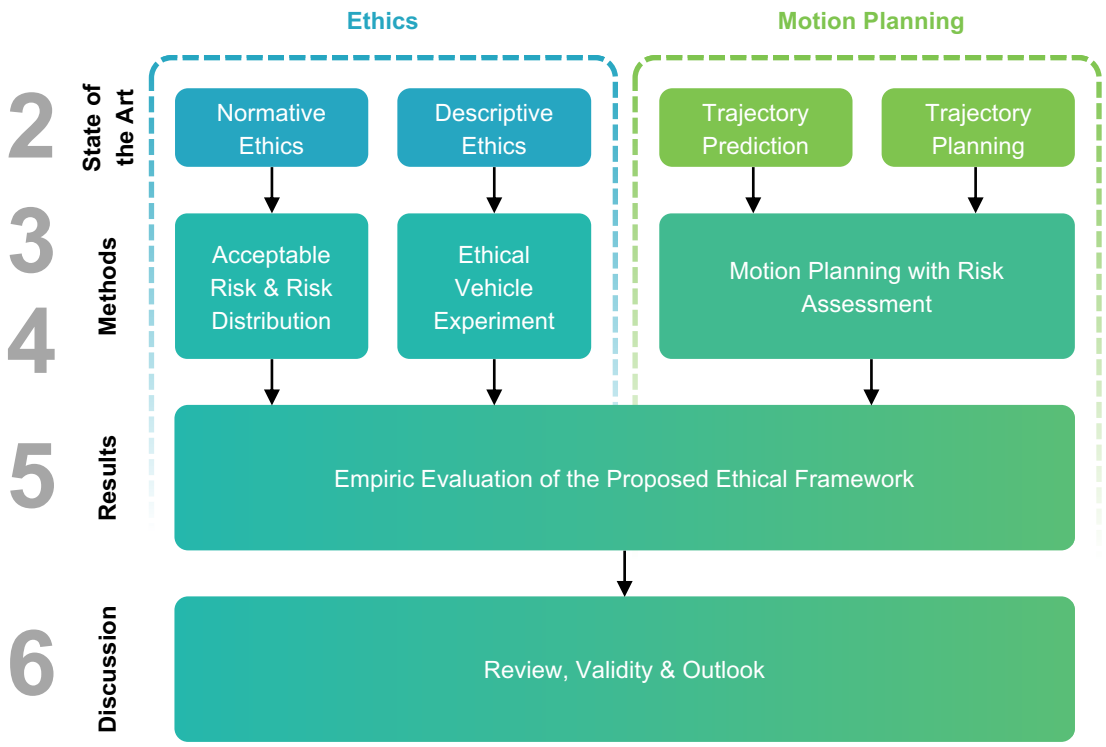


Figure 1.3: The schematic structure of the thesis shows the interdisciplinary characteristics of this work. Works from the fields of ethics, as well as AV motion planning, build the basis of this work and are consequently merged in the further course.

2 Related Work

This chapter lays the fundamentals upon which this thesis will build. The development of ethical algorithms for AVs involves an interdisciplinary approach, incorporating considerations from diverse fields such as computer science, engineering, philosophy, psychology, sociology, and law. The overall objective of the work in this area is to ensure that the AV technology is used in a way that aligns with human values and ethical principles.

To establish a common understanding of relevant terms, Section 2.1 provides definitions for key terms used throughout this work. Subsequent sections delve into the state of the art from two perspectives. Section 2.2 presents the related works on autonomous driving in the field of ethics. The challenge of decision-making is thereby initially delineated from further ethical challenges of AVs (Section 2.2.1). Additionally, works from the fields of normative ethics (Section 2.2.2) and descriptive ethics (Section 2.2.3) are presented in an organized manner, providing a reference for further development. Finally, the challenges from the ethical side for a decision-making algorithm are presented by the various recommendations of ethics committees and legislation in Section 2.2.4.

Moving forward, this chapter shifts the focus to the algorithmic side in Section 2.3. Ethical decision-making is represented in the software through motion planning, with trajectory planning (Section 2.3.1) and trajectory prediction (Section 2.3.2) as the pertinent components. As a result of the previous section, particular emphasis will be placed on approaches that include risk and uncertainty here.

To conclude this chapter, the research gap is identified based on the prepared literature and formulated in the form of research questions. The research questions are revisited in the discussion in Section 6.1, and their answers are critically reviewed.

2.1 Terms and Definitions

To establish a coherent foundation of comprehension, the following section defines the most relevant terms within the scope of this work. These definitions adhere to common usage and align with those found in the related literature. It is important to emphasize that in instances where a term may have multiple interpretations, only the specific definition provided here holds valid significance within the context of this work.

Autonomous Vehicle (AV): A vehicle that operates within a predefined Operational Design Domain (ODD) without human input. Consequently, this refers to SAE level 4 or 5 depending on the ODD [17]. Unlike Connected Automated Vehicles (CAVs), AVs are assumed to have no further communication, such as Vehicle-to-Vehicle (V2V) or Vehicle-to-Everything (V2X).

Trajectory Planning: Trajectory planning is a fundamental functionality as part of the AV software. Based on an environment representation and predicted trajectories of other road users, the task of trajectory planning is to calculate a collision-free trajectory that connects the vehicle's current state with a desired goal state.

Trajectory Prediction: As a functional part of the AV software, the task of trajectory prediction is to predict the trajectories of other road users based on the environmental perception of the AV without active communication.

Motion Planning: There are various definitions for motion planning available in the literature. In this work, motion planning describes the combination of trajectory prediction and trajectory planning.

Path: An ordered sequence of waypoints, which is intended to be followed by the AV.

Trajectory: An ordered and time-dependent sequence of waypoints. In contrast to a path, a trajectory includes a velocity profile.

Maneuver: A high-level characterization of the AV motion, regarding the position and velocity of the AV on the road [18].

Risk: This work defines risk in terms of collision risk for AVs. Consequently, risk is defined as the product of a collision probability and the severity in terms of personal harm as the result of a potential collision (Section 3.2).

Ethical Decision-making: Ethical decision-making is considered a process by which individuals would use their moral base to determine whether a certain issue is right or wrong [19].

Ethical Algorithm: An algorithm that actively considers aspects from ethical theories is denoted as an ethical algorithm in this work.

Fairness: There is no unique definition of fairness. Within this work, fairness is associated with the definition of John Rawls' Theory of Justice [20]. This approach will be discussed in Chapter 6.

2.2 Autonomous Driving Ethics

The field of AVs presents a number of ethical challenges, some of which are visually summarized in Figure 2.1. These challenges arise from various sources related to the advancement of technology. An overview of the variety of ethical challenges in autonomous driving is presented by [21, 22]. Some of these ethical issues directly influence the design and development of AVs, imposing specific requirements. Conversely, particular concerns, like liability [23] or implications to the labor market [24, 25], can be seen as largely independent of technological development.

Ethical issues related to technical development primarily stem from two sources. Firstly, AVs, due to their autonomy, make decisions that have moral implications. These decisions could possibly include the AV being forced to break traffic rules [26], for example. Moreover, these autonomous agents pose the risk of being hacked and abused [27]. Secondly, the increasing reliance on data-based techniques, particularly Machine Learning (ML) algorithms, introduces challenges in data handling and privacy [28]. The resultant models often operate as black-box systems, lacking explainability [29]. Additionally, model performance may be influenced by biases present in the training data, potentially leading to adverse consequences for underrepresented groups [30–32]. First OEMs pick up these challenges and discuss them in published reports [33].

While acknowledging the variety of ethical challenges with AVs, this work focuses on the decision-making of AVs. Again, there is a variety of challenges related to ethical decision-making, e.g., with regard to transparency and clear communication of the decisions made [34]. The focus of this work is on decision-making under risk and the associated fair distribution of risk. Nevertheless, other ethical challenges that are not the focus of this work, such as acceptance or transparency, will be considered as additional requirements in the development.

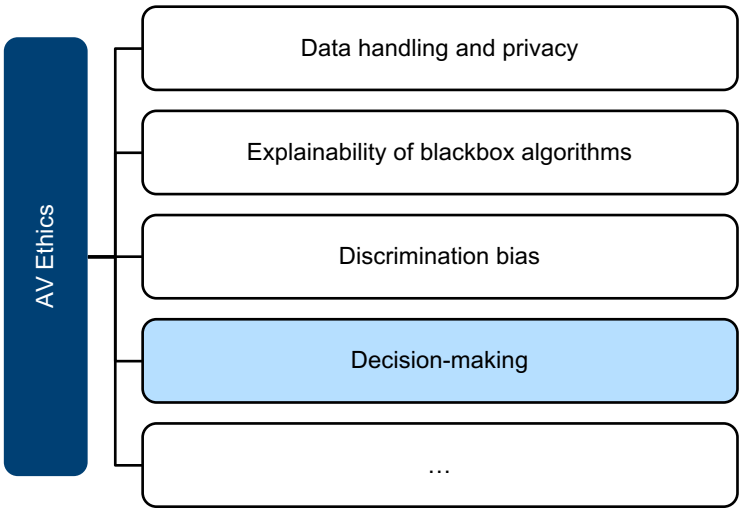


Figure 2.1: Overview of the ethical issues related to AVs. This work primarily concentrates on ethical decision-making within the broader ethical landscape.

2.2.1 Ethical Decision-making of AVs

The literature largely agrees that AVs will be involved in crashes and thus bring inherent risks into road traffic [35–37]. Previous failures of AVs in real-world testing that resulted in fatal crashes [3] underline this. OEMs also recognize that Vision Zero, meaning no traffic fatalities from AVs, is a utopia, and as long as human road users are involved, fatal accidents will occur [33]. There are various reasons for risk that cannot be mitigated completely [38]. For example, some objects are considered to be unpredictable [39], such as pedestrians, cyclists, or animals. Consequently, there is no way to avoid inevitable collision states in the current state of the art, but only to minimize the probability of entering one [40]. Moreover, the automotive standard for functional safety ISO 26262 [41] acknowledges that some inherent risk will remain in the system [42].

For this reason, AV's decision preceding crashes will have a moral component [4] as decisions with critical outcomes are put under the control of algorithms that were formerly the preserves of human beings. Encoding human morals in software is considered complex [4]. Studies revealed that in critical situations, human beings are panicked and act on their instincts [7]. In the case of AVs, however, such decisions must be programmed in advance. So, what used to be governed solely by reflexes and instinct with a human driver must be programmed by a deliberately designed algorithm in AVs.

The literature on ethical AV decision-making can be categorized according to their branches of ethics. Normative ethics, in general, aims to answer questions on how people should act. In contrast, descriptive ethics focuses on what people think is right and thus represents the current state as "is", whereas normative ethics rather focuses on the "ought". As a third category, meta-ethics concentrates on the subordinate questions, e.g., what "right" even means or who should provide answers to moral questions.

In terms of AV decision-making, questions beyond normative ethics arise, such as who should decide how AVs should behave [43–45]. Some works argue that each vehicle user should be allowed to choose the ethical view by which the vehicle should act [46]. Therefore, a so-called "ethical knob" is proposed with which the passenger can adapt the ethical settings [47]. Contrary studies found out using game theory that having mandatory settings would be in the interest of all road users [48]. This thesis does not intend to go deeper into this discussion or provide a solution to this issue. However, the disagreement that exists here must be taken into account in the development so that, ideally, no restrictions have to be made in this respect.

The Need for Ethics

In the literature, there are works that discuss whether and to what extent AV decision-making should be considered from an ethical perspective. Some works argue that the question of AV decision-making, especially when it comes to unavoidable accidents, must not be considered as a dilemma but from a pure safety perspective [49]. Consequently, ethical decision-making is considered a distraction with little practical value for the design of AVs [49]. Opposing arguments include that the safety of AVs is principally a question of ethics [42]. From a vehicle dynamics perspective, Davnall [50] argues that it is always the least risky to brake in a straight when a critical situation occurs, making ethical considerations obsolete. However, this strategy becomes less effective when the barriers in the fictive situation are not equidistant from the AV [51]. Consequently, prioritizing the shortest braking distance might not always yield the optimal solution. Evasion maneuvers have also been shown to be a better alternative to straight emergency braking [52]. Further arguments against ethical considerations for AV decision-making include that a substantial portion of the difficulties presented by AVs can be resolved by employing the same ethical decisions that humans have been making for centuries [53]. Therefore, it is assumed that there is little need to integrate ethics into AVs. Other suggestions are to refuse actively choosing, for example, in the case of dilemma situations, and decide instead randomly [54].

However, the majority of the works in this field argue for the importance of actively considering ethics. For example, studies have shown that ethical questions regarding AVs are important for potential users [55]. Not considering ethics is therefore seen to be a fundamental roadblock to mass adoption of AVs [56]. A large part of the literature and the public discourse in this area focuses on dilemma situations and, specifically, the trolley problem. Therefore, the applicability of the trolley problem to real AV applications will be highlighted in the following.

Debate on the Trolley Problem

Public discourse on AV ethics often focuses on moral dilemmas, such as the trolley problem [8, 57]. A simple and widely used version is as follows [51]:

"Imagine you are standing at a switch, and a trolley is speeding toward five people tied to the rails. Certainly, these people will definitely die if you do not intervene. There is the possibility of changing the switch. On the other rail, however, there is also a person tied up on the tracks who will surely die if the trolley takes this path. You can either do nothing, and five people will die, or you can pull the switch, and a single person will be killed. What will you do?"

Many articles have drawn comparisons between the ethical dilemmas posed by accidents involving AVs and the trolley problem [58, 59]. From the point of view of normative ethics, similar scenarios for AVs are formulated [60, 61], and possible solutions are proposed, e.g. based on criminal law [62]. However, the value of discussing trolley-like dilemma situations is discussed in the literature: Although there are works that underline the importance of trolley problems for the programming of AVs [63], there are various arguments against putting the focus on the trolley problem: Firstly, the outcomes of the trolley problem, namely the death of the victims, are postulated as certain events, which cannot be assumed in real-life scenarios [64]. Therefore, moral uncertainty must be part of the consideration [65]. Secondly, the trolley problem presents a binary choice while AVs possess the capability to navigate within a continuous solution space for trajectory planning. Figure 2.2 schematically shows that there is an infinite number of possible trajectories in a continuous action space when it comes to AV trajectory planning. Thirdly, it's important to note that depending on the version of the trolley problem, there may be a lack of crucial prior information regarding the circumstances leading up to the ethical dilemma. For example, the moral responsibility of the involved agents is argued to be important in such decisions [66, 67]. This information may be necessary to allow a morally well-founded decision to be

made. Other arguments discuss the relevance of the trolley problem in terms of the likelihood that it appears in real traffic [68]. In this course, solving the trolley dilemma could delay the rollout of socially useful AVs and, therefore, could have a negative impact [69].

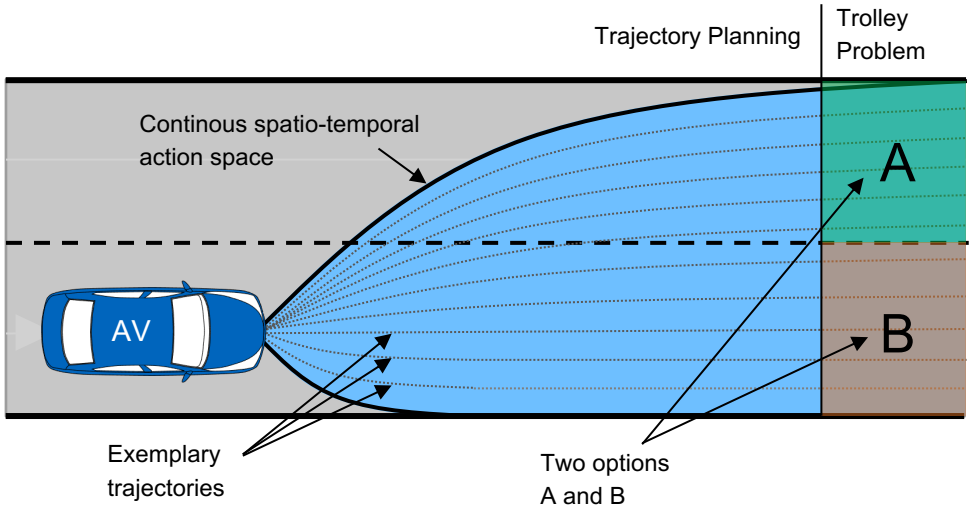


Figure 2.2: In contrast to the trolley problem, where a decision between two options (e.g., A and B) has to be made, in real AV trajectory planning, there are numerous trajectories as decision options.

The idea of having simple and fictive scenarios and confronting people with morally difficult questions is carried out by further works. In the sense of descriptive ethics, these works conduct user studies with trolley-like questions to better understand the desired behavior of potential AV users [70, 71]. These descriptive studies will be presented in more detail in Section 2.2.2.

The literature largely agrees that the trolley problem is an ill-suited benchmark [72]. Scholars recommend moving from trolley problems to questions of tradeoffs in values [73] and mundane traffic situation [68]. The core of the problem, namely finding a trade-off between various ethical principles, can be transferred to ordinary road traffic. Therefore, the trolley problem can be seen from a statistics point of view [14]. Even if the direct usability of the trolley problem is questionable, it can contribute to stress testing initial ideas about how self-driving cars should be programmed [74].

2.2.2 Normative Ethics

Various theories of normative ethics are suitable for providing guidance in AV decision-making. Ethical approaches to AV decision-making can be categorized according to their underlying ethical theory. For autonomous driving, four relevant ethical theories are frequently discussed in the literature: deontology, consequentialism, virtue ethics, and risk ethics. The following sections are structured to briefly explain the underlying theory and then present relevant works in the context of AVs with the strengths and weaknesses discussed. Few approaches in the literature combine multiple ethical theories, which are then listed in each corresponding section. Parts of this content overlap with the author's previous publication [51], which, however, focuses more on a requirements-based analysis of the theories.

Deontological Approaches

Deontological ethics is a class of ethical theories belonging to normative ethics. According to the deontological theory, an action is evaluated on the basis of whether it is in accordance with an obligatory rule and whether

it is committed on the basis of this obligation [75]. Thus, an action can be described as intrinsically good or bad, regardless of its consequences.

In the AV context, these obligations can be formulated as rules, such as Asimov's three laws of robotics [76]. In hierarchical order, these laws represent a prioritization in which the highest priority is that robots must not harm humans. For AVs, a rule-based approach seems straightforward, as road traffic regulations are also formulated using rules. In general, such rule-based ethical theories may represent promising application possibilities for AV decision-making since they offer a computational structure for judgment and, thus, at least from a practical perspective, seem achievable [77]. The approach of a Responsibility Safety Shield (RSS) [78] sets additional rules that should prevent AVs from accidents by setting responsibilities. However, following the strict rules from RSS was shown to lead to undesirable consequences [79], as, for example, obligations must not be fixed as they can change over time. Moral decisions and obligations are not absolute but depend on the context [10]. Regarding critical situations, an assumed priority order of objects to crash with was implemented as part of a Model Predictive Control (MPC) [80]. Thornton et al. [81] extended a similar MPC approach with ethical considerations, where deontology can be used to set the MPC's constraints. They combine this with deriving cost functions as part of a consequentialist approach, concluding that deontology is insufficient to use alone. It was shown that the prescriptive nature of setting rules might end up in situations where AVs cannot move anymore, e.g., if they would obey the law [82]. Censi et al. [83] aimed to mitigate that problem by presenting a large hierarchical rulebook where higher priorities can revoke rules. While their implementation consisted of 15 exemplary rules, they estimate about 200 rules to be needed to cover a city like Singapore with a level 4 system. In this regard, Lindner et al. [84] underline the computational complexity that must be considered for implementing ethical principles. Other approaches formulated eight rules, such as "Do not harm people outside the vehicle" or "Do not harm passengers" and prioritize them according to different moral views [85]. For an unmanned aircraft, it could be proved formally that the prototype only performs an unethical action if the rest of the actions available to it are even less ethical [86]. Despite the vast amount of necessary rules, the literature argues that such rule-based approaches may ignore context-specific information [87], such as the probability of occurrence of current and future conditions. Hence, the AV could perform dangerous behaviors to comply with its strict rules [88].

Another group of ethical approaches to AV decision-making is based on contractualism as part of deontology. Contractualism goes back to Rawls [20] and promotes the underlying idea of seeking principles that individuals in a social contract would agree to. Using this idea to address the normative challenges of autonomous driving could reflect the plurality of moral doctrines within society [89]. In addition, this approach is not limited to specific or hazardous situations but can be applied to any algorithmic AV decision [90]. As a result of the idea of a social contract, Rawls promotes the maximin principle for decision-making as a societal consensus. This principle encourages decision for the option in which the worst possible outcome (max) is minimized (min). On these grounds, Leben [91] developed a similar rule for AV decision-making. This rule compares survival probabilities and selects the alternative that assigns the greatest survival probability to the worst-off person. Criticism of this approach argues that it gives undue weight to the worst-off and it neglects a second dimension of collision probability by only considering a survival probability as a measure of expected harm [92].

Consequentialist Approaches

The concept of consequentialism covers ethical theories that judge the moral value of an action on the basis of its consequences [93]. A prominent variant of consequentialism is utilitarianism, which was originally formulated by Jeremy Bentham. Utilitarianism advocates for the maximization of human well-being. This ethical theory evaluates the moral rightness of action exclusively by examining its outcomes [94], with the aim of maximizing the anticipated overall utility.

For decision-making in autonomous driving, optimized outcomes can be achieved, for example, by minimizing the resulting harm caused by AVs. Therefore, cost functions of behavior and trajectory planning algorithms are suitable for implementing utilitarian principles [81] by selecting the choice with the lowest cost [5], which could be, e.g., the number of victims in car crashes [95]. To assess the consequences of AV decisions, utilitarian approaches require estimation models that connect various decision options with their foreseeable consequences. To estimate the expected harm for various options, a lumped mass-spring model of an AV colliding into an immovable rigid barrier was used [96]. Kumfer et al. [97] showed in a simulation of three scenarios that a utilitarian cost function provides the greatest net benefits to society. However, they conclude that some drivers may reject being potentially sacrificed to protect other drivers. Moreover, the German ethics commission [98], for example, considers this inadmissible. Further limitations are that a utilitarian approach fails to account for the fundamental difference between those involved and uninvolved in an impending crash [99]. To approximate a compromisable approach, scholars advocate the combination of deontological ethics and utilitarianism in the form of a relative weighing of costs and options [100]. More advanced approaches extend the basic utilitarian idea to include probabilistic considerations and thus better reflect reality [101]. Therefore, the utilitarian principle is a distribution principle in risk ethics, which will be presented later.

Approaches from Virtue Ethics

Virtue ethics, with its roots tracing back to the philosophical works of Plato and Aristotle, frames morality as a matter of character. In this ethical perspective, the embodiment of virtues stands at the core of a well-lived life. Specifically, for the human condition, the cardinal virtues encompass prudence, courage, temperance, and justice [94].

Cognitive machines, such as AVs, should analogously exhibit such virtues [102] considering the kind of person or organization the AV represents. An ambulance may follow different virtues than a taxi, as it would be acceptable to run a red light, for example [82]. Thornton et al. [81, 103] refer back to virtue ethics to determine parameters in their models according to the role of the vehicle. Other approaches state that virtues within machines cannot be pre-programmed but are the results of machine learning [102]. The technical feasibility is shown by imitation learning for real-world driving [104] or Reinforcement Learning (RL) [105]. In this process, ethics in the form of a set of virtues can provide guidance as a pattern of positive signals [106]. Further implications of virtue ethics regarding AV safety rely on engineering judgment [42]. Accordingly, the software engineer should be guided by the question: "Do I want to be known as the person who programmed the vehicle to do the following?" Virtue ethics has only a few concrete references to real implementation. Therefore, it remains unclear how exactly the virtues of AVs should be programmed.

Approaches from Ethics of Risk

The subject of risk ethics is the moral ex-ante evaluation of actions whose consequences are fraught with uncertainty with regard to their occurrence, benefit, and harm [107]. This theory deals with the general question of the conditions under which a person may expose himself or others to risk. While discussion often focuses on life-or-death dilemmas, they may be rare in practice. Therefore, four far more common examples of routine driving that require decisions with some level of ethical reasoning about how to distribute risk are proposed [108]. These scenarios include, for example, choosing a following distance or a lateral position as in Figure 1.2.

There are various definitions of risk in the literature. For example, risk can describe an unwanted event, the probability of an unwanted event, or a statistical expected value as the product of probability and some measure of its severity [109]. Recent work has underlined the importance of considering risk and uncertainties

in AV decision-making [110]. Hence, risk ethics offers a promising opportunity, as it is already being used in other applications, such as organ donation [111] or radiation exposure [112]. Risk ethics expands upon established decision principles by incorporating uncertainty considerations to inform decision-making in uncertain situations. For example, the Bayes principle is according to a utilitarian approach and strives to maximize the expected utility. As part of a crash optimization algorithm, tree sorting on survival probabilities can be used [113]. On the ground of contractual ethics, further works suggest always aiming to maximize the lowest occurring survival probability [91]. Like utilitarianism, this requires risk estimation models according to the risk measure used. Risk measures are also used in functional safety, e.g., in ISO 26262 [41], which could serve as a starting point to define agreed risk levels for AVs [114]. Based on expected harms, the Ethical Valence Theory (EVT) provides a decision strategy based on the moral claims of various road users and their mitigation [115].

2.2.3 Descriptive Ethics

In contrast to normative ethics, works in the field of descriptive ethics study people's views on questions of morality. In recent years, many studies have been conducted that have shed light on various aspects of human decision-making in autonomous driving. Most of these studies intend to support the formulation of computational models [116, 117]. Therefore, having a comprehensive understanding of the existing moral judgment theories is considered crucial to programming realistic and accountable AV behavior [117]. Table 2.1 provides an overview of the existing work in this field, paying special attention to the focus of the study.

With more than 40 million participants from 233 countries, the Moral Machine Experiment [70] was the largest study here. It aimed to reveal cultural differences in decisions where an AV has two options to crash into. The options differed by 13 characteristics, such as the age, gender, or social status of the people. Figure 2.3 shows an exemplary question of the experiment, which includes different genders, sizes, or ages of the humans involved. Another component is whether to swerve and thus actively intervene or rather stay on track. The Moral Machine Experiment tightened a public discussion about which criteria these moral decisions should be made on. The majority of the following works, as well as the legislation [13] and recommendations from the ethics committees [98], come to the conclusion that the factors used in the study (e.g., gender, age, social status) are inadmissible. The basic concept of the study, which weighs human lives against each other, is also subject to criticism [118].

Researchers identified two main impact factors in human decision-making [119]: Firstly, decisions depend on the personal perspective of those being asked [120]. Secondly, they depend on the amount of time in which an answer must be given. In critical traffic situations, humans are expected to decide in a fraction of a second [121]. These decisions are dominated by panic and instinct [7], and the less time humans have, the less consistent their decisions are [122]. Under time pressure, people fail to decide utilitarian as they typically would with more time [123].

Further studies investigate users' preferences as to whether the AV should perform decisions based on utilitarianism or egoism in critical situations. Here, the studies come to ambiguous results. Bonnefon et al. [15] conclude that although users agree on utilitarian decision-making as the most moral, they prefer self-protective vehicles for their own, which is confirmed by the work of Liu et al. [124]. Other studies suggest that drivers would prefer utilitarian decisions even if they were adversely affected themselves [71, 125, 126]. However, instead of options with fixed outcomes, they use probabilities of survival as risk measures to perform the decision on. Hence, the perceived risk seems to have an impact on how human beings tend to decide [127]. The importance of investigating moral judgment and decision-making in situations in which consequences are only probabilistically known or not precisely quantified is highlighted in further studies [128]. Therefore, Table 2.1 reviews the discussed works against the background if they incorporate risk or uncertainty. As can be seen, only a few works take this seemingly important component into account.

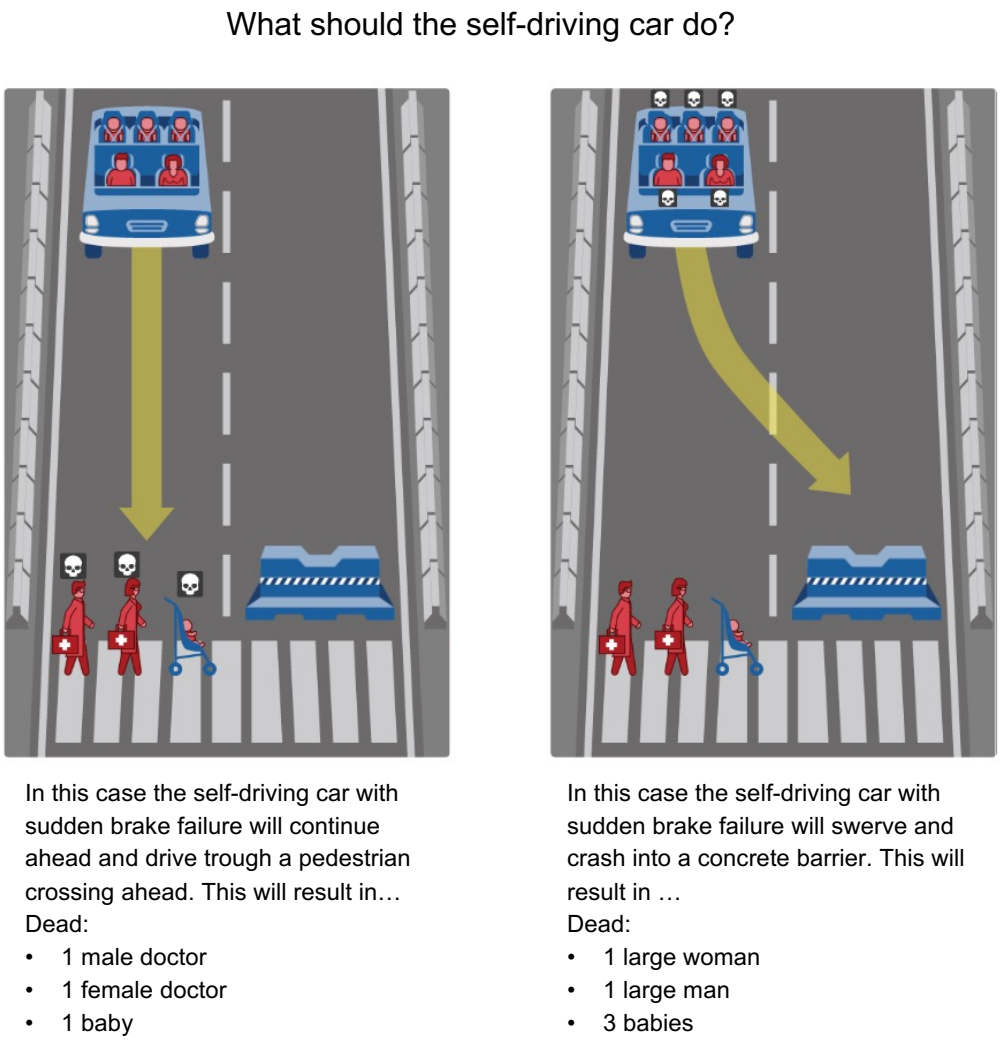


Figure 2.3: Exemplary question from the Moral Machine Experiment [70], which can be accessed at https://www.moralmachine.net/. The experiment focuses on decision-making based on 13 features, such as gender, age, or the number of people or animals that will die.

Moreover, the lack of real-world validity in the simplified scenarios acted by the studies is also criticized [129]. Grasso et al. [130] found different human behaviors depending on the level of immersion in the study. For this reason, driving simulators are used to investigate instinctive decisions next to reflected decisions with more time [131, 132].

The findings of all these studies aim to shed light on how the AV should behave. Whether these findings from descriptive ethics can be transferred to a normative decision criterion is a matter of discussion [133]. However, determining ethical decision-making in a participatory manner is considered to require involving all those who will be affected [134]. In that case, such ethical preferences, as a result of participative studies, must be screened for bias and only be deployed to the degree to which they match major ethical theories [135].

2.2.4 Legislation & Ethical Guidelines

In recent years, government sectors, as well as industries, have provided ethical guidelines to ensure that emerging technologies in relation to Artificial Intelligence (AI) benefit humanity as a whole [137]. They are particularly important as they establish the framework and often are intended as a template for legislation. Therefore, the following section will provide an overview of the current guidelines of various institutions with a focus on Germany and Europe.

The EU set up an expert group in 2019 that declared trustworthy AI as the leading goal [138]. Accordingly, trustworthy AI has three components that an AI system must meet throughout its life cycle: it should be (1) lawful, (2) ethical, and (3) robust both from a technical and social point of view. The requirement of adhering to ethical principles is further broken down into four essential principles that must be followed: respect for human autonomy, prevention of harm, fairness, and explicability [138]. In addition, the report promotes paying particular attention to vulnerable groups, such as children, for example. The experts acknowledge that AI systems may bring risks and have negative impacts. Therefore, developers should adopt adequate risk measures to mitigate these risks when appropriate and proportionately to the magnitude of the risk. This risk-based approach is later adopted by the EU AI act that sets out harmonized rules for the development and use of AI in the European Union [139]. This first regulation on AI assigns applications to three risk categories. First, applications and systems that pose an unacceptable risk, such as social scoring systems [140], for example, are prohibited. Second, high-risk applications, such as scanning tools using computer vision that rank job applicants [141], are subject to specific legal requirements. Third, applications that are not explicitly banned or classified as high-risk remain largely unregulated.

Jobin et al. [142] analyzed 84 reports on AI ethics worldwide, which also include reports from the private sector from companies, such as Google [143] or IBM [144]. They found a convergence around core principles to follow, such as transparency and fairness, for example. However, they also conclude that these reports lack clear instructions on practical implementations. Therefore, further research is identified as being necessary.

Ethical recommendations have been further detailed for the application of autonomous driving. In 2017, the German Federal Ministry for Transportation and Infrastructure ordered a group of experts to provide the "necessary ethical guidelines for automated and connected driving" [12, 98]. This led to 20 guidelines that can be organized into five categories: "unavoidable accident situations", "data security and data economics", "human-machine interface", "responsibility for software and infrastructure", and "ethical context beyond traffic" [98]. In terms of AV decision-making algorithms, the guidelines regarding unavoidable accidents are of particular interest. For the development of an ethical AV algorithm, the following four points of the ethics committee are crucial:

Table 2.1: Overview of recent works that conduct studies to describe human morality according to descriptive ethics. The requirement of reflecting uncertainties is only covered by a few works.
[1]: Studies consisting of multiple stages have varying numbers of participants.

Authors	**No. of participants**	**Incorporating uncertainty/risk**	**Focus of investigation**
Frison et al. 2016 [71]	40	Yes	Many people would risk their own life to save others according to a utilitarian view
Bonnefon et al. 2016 [15]	182 - 451[1]	No	Although agreement on utilitarian as the most moral, users prefer self-protective models according to a selfish view
Wintersberger et al. 2017 [125]	40	Yes	Impact of personal perspective whether the user's position in the experiment is known or not known
Suetfeld et al. 2017 [122]	105	No	Time pressure decreases consistency in human decision-making
Awad et al. 2018 [70]	40 Mio.	No	Cultural differences in decision-making
Bergmann et al. 2018 [126]	189	No	Influence of age, as well as egoism and altruism
Faulhaber et al. 2019 [7]	189	No	Human decisions in moral dilemmas are largely described by utilitarianism
Frank et al. 2019 [119]	807	No	Two human biases in decision-making: Time taken for the decision (instinctive or reflected) and personal perspective
Meder et al. 2019 [128]	1638	Yes	Importance of investigating moral judgements under risk and uncertainty
Samuel et al. 2020 [123]	32	No	Differences between instinctive and reflected human decisions in a driving simulator
Lucifora et al. 2020 [131]	84	No	Differences between instinctive and reflected human decisions in a driving simulator
Grasso et al. 2020 [130]	84	No	Impact of immersion of the study, e.g., abstracted scenario or driving simulator
Lucifora et al. 2021 [132]	100	Yes	Differences between instinctive and reflected human decisions in a driving simulator
de Melo et al. 2021 [127]	94 - 276[1]	Yes	Utilitarian choices dominant but dependent on perceived risk
Mayer et al. 2021 [120]	325	No	Impact of road user type (e.g., pedestrian or passenger)
Liu et al. 2021 [124]	580	No	Selfish AVs are more accepted than utilitarian
Shigeharu et al. 2023 [136]	683	No	Cultural differences in decision-making

R1: The system must be designed in a way that dilemma situations do not appear

R2: Higher priority of persons over animal or property damage

R3: Human lives must not be counted against each other and bystanders must not be sacrificed

R4: Qualification according to personal characteristics (such as age and gender) is prohibited

German ethics comittee [12]

However, although the commission is clear that human lives must not be compensated for (R3), it is remarkable that they agree with balancing risks against one another [98].

In a similar way, the European Commission ordered an expert group to work on the "Ethics of Connected and Automated Vehicles" [11]. The resulting 20 recommendations fall into three areas: (1) road safety, risk, and dilemmas, (2) data and algorithm ethics, and (3) responsibility. For AV decision-making, the following recommendations are considered relevant:

R5: Redress inequalities in vulnerability among road users (#5)

R6: Manage dilemmas by principles of risk distribution and shared ethical principles (#6)

R7: Enable user choice, seek informed consent options and develop related best practice industry standards (#8)

R8: Promote a culture of responsibility with respect to the obligations associated with CAVs (#17)

EU Horizon expert group [11]

These recommendations found their way into Level 4 automated driving legislation in the EU [13], which is based on a similar German legislation [145]. The regulation now in force in the EU provides for four further requirements in relation to ethical decision-making by AVs:

R9: The Autonomous Driving System (ADS) shall be able to detect the risk of collision with other road users.

R10: In the event of an unavoidable alternative risk to human life, the ADS shall not provide for any weighting on the basis of personal characteristics of humans.

R11: The protection of other human life outside the fully automated vehicle shall not be subordinated to the protection of human life inside the fully automated vehicle.

R12: The vulnerability of road users involved should be taken into account by the avoidance/mitigation strategy.

European Legislation [13]

In summary, the investigation of the legal landscape in terms of AV ethics yielded three relevant sources that provide requirements for AV development. These twelve requirements provide instructions but lack concrete implementations so far. Consequently, these requirements will be considered in this work to build upon.

2.3 Motion Planning

There are two essentially different approaches to the AV software architecture. First, in end-to-end approaches, (usually learning) algorithms are developed that directly calculate actuator signals from sensor input. An overview of these approaches is presented by [146, 147]. End-to-end approaches to the AV software architecture benefit from joint feature optimization for perception and planning [148]. A major challenge of these approaches, which has hindered their widespread use so far, is the explainability and interpretability of these models [149]. The second and more common approach today provides for a modular subdivision of the software according to different functionalities. Following the *Sense-Plan-Act* approach from robotics [150], the essential software components *Perception*, *Planning*, and *Control* are derived as in Figure 2.4 [151, 152]. *Prediction* can be treated as part of perception or planning or as a separate intermediate step, as in Figure 2.4.

Since the lack of explicability of end-to-end approaches is particularly reflected in decision-making, the work of this thesis is based on the modular approach. This also allows a targeted view of the decision-making process decoupled from pre-scheduled functionalities from Perception. Therefore, the focus of this work is placed on the motion planning of AVs, assuming a given environment model as perception output.

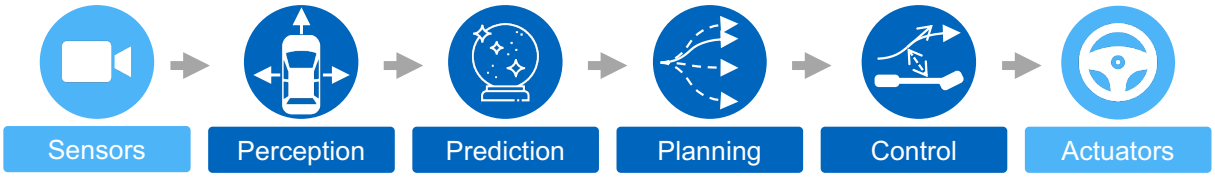


Figure 2.4: High-level overview of the modular approach to the AV software architecture [153]. Based on the sensor signals, the perception module provides an environment representation. The dynamic objects are predicted for their future behavior, and a collision-free trajectory is planned. The control module uses this trajectory to calculate steering signals for the actuators.

The task of motion planning in AV software is to calculate a trajectory based on an environment model from perception, which can then be converted into actuator commands by a controller (Figure 2.4). Figure 2.5 shows the three essential components of motion planning: First, a global path is an ordered collection of waypoints and provides guidance to a goal state. However, it does not provide a velocity profile and does not consider potential collisions, as dynamic objects are not considered. The second component, trajectory prediction, predicts the future trajectory of all relevant detected objects. On the basis of these predictions, trajectory planning aims to calculate a collision-free trajectory as the third component.

Predicting the trajectories of surrounding road users as underlying uncertain assumptions for trajectory planning plays a decisive role in motion planning. Therefore, this section will first reflect on the state of the art in trajectory planning before going into detail on trajectory prediction. Since the state of the art in ethics, as well as the previously presented legislation, motivates a risk-based approach, particular emphasis will be placed on this. A special focus will be placed on risk-aware trajectory planning and probabilistic trajectory prediction.

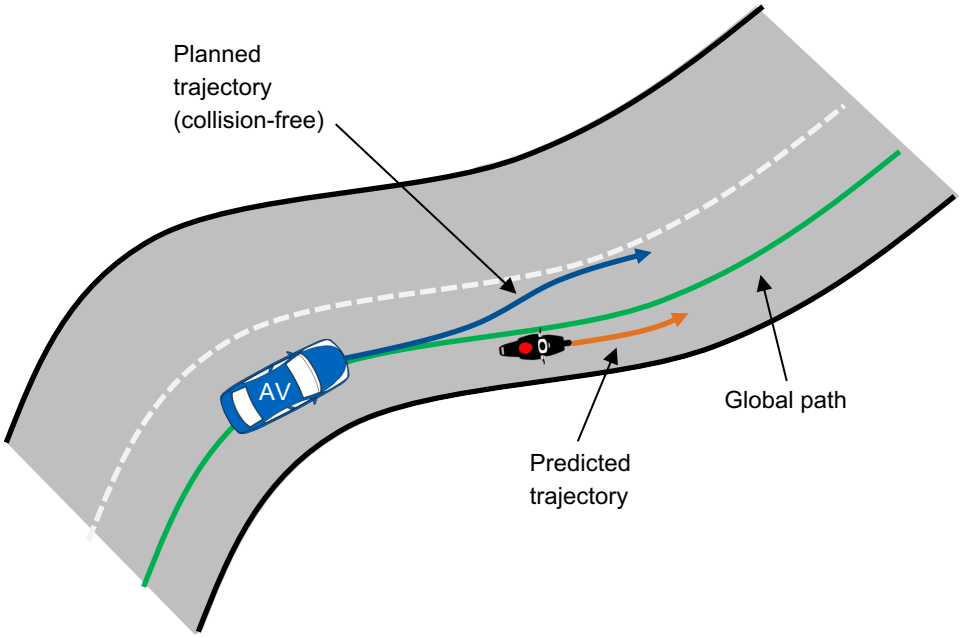


Figure 2.5: Schematic illustration of the essential motion planning components. Given an environment representation, a global path provides a connection of waypoints to the desired goal state. Based on the predicted trajectories of dynamic objects and the global path, a collision-free trajectory is planned.

2.3.1 Trajectory Planning

As a part of motion planning, trajectory planning describes the process of finding a collision-free trajectory that connects the vehicle's current state with a desired goal state. Some approaches from the literature subdivide trajectory planning. Accordingly, these approaches perform intermediate steps in so-called Behavior Planning [154, 155], Maneuver Planning [18, 156], or Mission Planning [157]. All follow the same idea to narrow down the possible trajectories in a first step (e.g., to a maneuver) before the final trajectory is selected. As most of the approaches presented in the following do not require this intermediate, the focus will be on the trajectory planning itself.

Most approaches to trajectory planning can be considered as consisting of two parts: one part that describes the planning objective (e.g., using a cost function) and a second part that calculates a trajectory according to the given objective [158]. In this regard, the next paragraph will present various approaches to the planning objective to discuss later how a suitable trajectory according to this objective can be found.

Planning objective

In general, the objective of a planned trajectory is described using cost functions and constraints [159]: Constraints include, for example, that the AV must reach a goal region $\mathcal{G}_S \subset \mathbb{R}^n$ without causing a collision with obstacles. Thus, the occupancy of the AV must be in the free space $\mathcal{W}_{S,\text{free}}(t) \subset \mathbb{R}^2$ for all $t \in [t_0, t_f]$, where t_0 is the initial timestep and t_f is the final timestep of the planning task. The trajectory should not only be free from collisions but also physically feasible. Therefore, the evaluation according to physical constraints is performed using a dedicated vehicle model, such as a point-mass, single-track, or double-track model. Additionally, further restrictions must be considered, such as traffic rules [160] for example.

Besides the hard constraints, the cost function incorporates objectives that express desirable conditions that may not be strictly mandatory. The cost function is a mathematical representation of these objectives that should be optimized as part of the trajectory planning process. Since there are multiple objectives to consider during trajectory planning and objectives may differ, e.g., according to the use case of the AV, there are

various cost terms used in the literature. Those cost terms can be assigned to four different superordinate categories of objectives [159]: safety, mobility, comfort, and driveability.

Safety factors may include minimizing the risk of collision or ensuring that the vehicle is within the legal speed limit. Mobility factors revolve around enhancing the efficiency and effectiveness of the vehicle during travel. This includes minimizing travel time to reach destinations promptly and reducing energy consumption to improve overall sustainability. Comfort factors are focused on providing a pleasant experience for passengers. It involves minimizing abrupt changes in acceleration, also known as jerk, and maintaining a smooth and stable ride. Driveability factors focus on ensuring that the vehicle can navigate the road seamlessly. This includes designing the vehicle to follow the curvature of the road naturally and handle sharp turns.

To combine different objectives, cost functions are usually designed as a weighted sum of various cost terms [161, 162]. Equation 2.1 describes the total costs J_{total} that a trajectory u over time t is associated with consisting of multiple cost terms J_i and their corresponding weight parameters w_i.

$$J_{\text{total}}(u(t)) = \sum w_i J_i(u(t)) \tag{2.1}$$

Table 2.2 provides an overview of various cost terms J_i that are commonly used in the literature [161–164]. As a current topic of research, this overview does not claim completeness, and further formulations for similar objectives are conceivable.

Table 2.2: Commonly used cost terms in AV trajectory planning. A more comprehensive overview can be found in [159].

Cost term	**Formula**
Acceleration	$J_A = \int_{t_0}^{t_f} a^2 \mathrm{d}t$
Distance to obstacles	$J_D = \int_{t_0}^{t_f} \max(\xi_1, ..., \xi_N) \mathrm{d}t$
Energy	$J_{EN} = \int_{t_0}^{t_f} P(x,u)^2 \mathrm{d}t$
Inverse duration	$J_{ID} = \frac{1.0}{\tau}$
Jerk	$J_J = \int_{t_0}^{t_f} \dot{a}^2 \mathrm{d}t$
Lane center offset	$J_{LC} = \int_{t_0}^{t_f} d^2 \mathrm{d}t$
Orientation offset	$J_O = \int_{t_0}^{t_f} (\theta_{\text{des}}(x(t)) - \theta(t))^2 \mathrm{d}t$
Path length	$J_L = \int_{t_0}^{t_f} v \mathrm{d}t$
Steering angle	$J_{SA} = \int_{t_0}^{t_f} \delta^2 \mathrm{d}t$
Steering rate	$J_{SR} = \int_{t_0}^{t_f} a_\delta^2 \mathrm{d}t$
Velocity offset	$J_V = \int_{t_0}^{t_f} (v_{\text{des}}(x(t)) - v(t))^2 \mathrm{d}t$
Yaw rate	$J_Y = \int_{t_0}^{t_f} \dot{\Psi}^2 \mathrm{d}t$

In addition to minimizing the distance to dynamic obstacles, there are further, more sophisticated safety measures that can be used in the cost function. Table 2.3 provides a noncomplete overview of such safety measures, some of which are used as cost functions for trajectory planning, such as Time-to-Collision (TTC), for example. One approach behind these measures is to describe how close or likely a collision with a dynamic obstacle is so that the planning algorithm prefers a state where a collision is unlikely. This applies for TTC, Time Exposed Time-to-Collision (TET), or Time-to-React (TTR), for example. Other approaches seek to quantify the severity of a potential collision. For example, Crash Index (CI) uses a kinetic energy model to provide a standardized index of the severity of a collision.

Table 2.3: Safety metrics for road traffic, which can be adapted to the motion planning of AVs. Further criticality measures can be found in [173].

Risk metric	Definition
Time-to-Collision (TTC)	Time until a collision will occur between two vehicles if the collision course and speed difference are maintained [165]
Time Exposed Time-to-Collision (TET)	Summation of all timesteps (over the considered time period) that an AV approaches a front vehicle with a TTC-value below the threshold value TTC* [166]
Time-to-React (TTR)	Time which is left to avoid the collision within the physical constraints of the vehicle [167, 168]
Crash Index (CI)	Influence of speed on kinetic energy involved in collisions
Aggregated Crash Index (ACI)	Quantitative measure to reflect the collision risk using the accommodability of freeway traffic state to a traffic disturbance [169]
Deceleration Rate to Avoid Crash (DRAC)	Minimum deceleration rate required by the following vehicle to avoid a crash [170]
Collision probability	Calculation of collision probabilities using Monte-Carlo-sampling [171] or Stochastic Reachable Sets [172]

As one of the most commonly used metrics, the EU Regulation [13] uses TTC for the technical specifications for the type approval of AVs. Therefore, the regulation provides further definitions and threshold values for the TTC for various scenarios. For example, for a cut-in scenario with standing or unfastened vehicle occupants with a relative velocity of $20\ km/h$, a TTC of $1.32\ s$ shall be maintained [13]. However, how such requirements should be implemented into software remains an open topic.

Although most of the literature uses a fixed weighted cost function for trajectory planning, there are works that find it useful to adapt these weights dynamically to different situations, for example, using Inverse Reinforcement Learning (IRL) [174, 175]. Other, more recent approaches use ML for the whole planning task. An overview of these approaches is presented by [176]. With the first available datasets, such as nuPlan [177], these approaches have gained importance. However, since they lack similar disadvantages as the End-to-End approaches discussed above, these methods will not be focused on here.

With given constraints and a cost function as the planning objective, the trajectory planning task is to find a trajectory u that minimizes the cost function $J_{\text{total}}(u)$ with respect to the given constraints. The literature provides different approaches to finding the best trajectory according to these requirements. The underlying methods are usually categorized into graph-based, sampling-based, and optimization-based [178–180]. In light of the context of this thesis, probabilistic and risk-aware approaches are presented as an additional category. The following will discuss these methods with respect to their advantages and disadvantages for real-time application in AVs.

Graph-based approaches

Graph-based approaches to trajectory planning are characterized by a discretization of the AV's configuration space as a graph. The graph consists of vertices that represent a finite collection of vehicle configurations and edges as transitions between vertices [181]. Selection of the final trajectory is performed by means of a cost function, such that the costs of vertices are minimized. Current algorithms differ in strategies for graph generation and methods for finding the optimal trajectory [179]. Typical geometric methods for graph generation are realized by creating a lattice or tree of motion primitives as in [182–184]. Commonly used strategies for finding the optimal trajectory within a graph are Dijkstra [185], A* [186], and D* [187]. The main advantage of graph-based motion planning is that it mitigates the problem of convergence to local minima by performing a global search in the configuration space. Depending on the discretization, the computational

effort can become challenging, so recent works reduce the dimension and perform local path planning using a graph and a separate velocity planning as the second step [181]. Further methods create the graph offline so that during runtime, the graph must only be evaluated [188], and the velocity profile can be considered here. However, due to the pre-computation, this is only feasible for small and known environments, such as race tracks [189].

Sampling-based approaches

The idea behind sampling-based approaches to trajectory planning is to solve the optimization problem by first generating a large set of trajectories and then evaluating the generated trajectories with respect to the constraints and cost function in a second step. Thus, these approaches do not enforce a specific representation of the free configuration set and dynamic constraints [179]. Along a reference path, sampling-based methods generate paths or trajectories according to a sampling scheme. A common approach on which many of the algorithms build up is Rapidly Exploring Random Trees (RRTs) [190, 191]. For example, the deployment in a real vehicle is shown by [192, 193]. With these approaches, such as RRT, optimal solutions are not given [191]. However, more sophisticated methods based on this have proven asymptotic optimality [191, 194]. The sampling scheme can be probabilistic or deterministic. For probabilistic roadmaps (PRMs) [195, 196], the advantages of deterministic sampling over random sampling have been shown in theory and practice [197]. For a reference path, the centerline of the AV is frequently used to guide the sampling process. Given the reference path, the sampling space can be of varying dimensions. After a sampling of spatial points, the velocity profile can be generated in the next step [198]. Expanding the sampling space into the temporal dimension allows assessing final trajectories in a single step [199]. Further variables, such as vehicle orientation or planning horizon, can also be added to the sampling space [199]. Werling et al. [200, 201] sample along the lateral distance to a reference path, target velocities, and time horizons. They split trajectory generation into lateral and longitudinal movement, utilizing 4th and 5th-order polynomials to ensure jerk-optimal trajectories.

Optimization-based approaches

Optimization-based or variational methods for motion planning do not discretize the configuration space of the vehicle. The problem of trajectory planning is formulated as an optimization problem. A common approach is to solve a constrained optimization problem within a receding horizon control [202, 203]. Recently, nonlinear continuous optimization techniques have been used to find the trajectory that minimizes a given cost function. Schwarting et al. [204] compute collision-free trajectories with a nonlinear MPC. Subsoits et al. [205] formulate the trajectory optimization problem as a quadratically constrained quadratic program and thus obtain solution times of less than $20\ ms$ even with a $10\ s$ planning horizon. The progressive improvement in the area of nonlinear solvers enables more and more competitive variational algorithms. The main advantage of constrained optimization is the smoothness of trajectories and direct encoding of the vehicle model in trajectory planning [178]. However, if the formulated optimization is not convex, it converges only to local minima. Therefore, these approaches are suitable for the re-optimization of a given trajectory or path [206]. For example, combining an optimization-based algorithm, such as MPC, with a graph-based step to ensure a convex optimization problem has proven to be a successful method [207].

Probabilistic and risk-aware approaches

Most of the methods mentioned earlier presume an environment representation devoid of uncertainties. However, in road traffic, uncertainties arise due to factors like imperfect perception, occlusions, and limited sensor range [208], which motion planning algorithms can actively address. Apart from integrating risk

metrics or uncertainties into the cost function to define the planning objective, there exist approaches capable of accommodating uncertainties. Analyzing the sensitivity of motion planning uncertainties reveals the importance of different uncertainties depending on the scenario [209]. This understanding allows defining uncertainty bounds to prevent unacceptable large deviations in the planned trajectory [210]. Collision probabilities are taken into account by navigation algorithms for robotics [211] using Monte Carlo-based methods [212]. In contrast to approaches that solely consider velocities for calculating collision costs [213], an internal energy model is employed to gauge the severity of a collision [211]. Introducing existence probabilities of relevant objects has shown to be beneficial to tolerate faulty detections of phantom objects [214]. Another important reason for uncertainty that is focused on is due to occluded areas [215, 216]. A strategy for handling uncertainties involves Partially Observable Markov Decision Processs (POMDPs). Recent work here considers uncertain predictions of other road users and benefits from the continuous execution of gathering more information [217], leading to a reduction in general uncertainty. However, this method lacks comprehensive trajectory planning, focusing solely on longitudinal accelerations and excluding lateral planning.

In the pursuit of real-time feasibility while maintaining a broad action space, additional assumptions are introduced, including the neglect of interactions between agents [208]. Research explores interactions with human-driven vehicles and utilizes a probabilistic model to emulate human behavior [218]. A conditional value-at-risk objective function [219] is adopted to quantify an upper limit on the divergence from the probabilistic human model. Similarly, another approach integrates a risk model based on collision probability into trajectory planning, tuning parameters to simulate human-like behavior [220]. An alternative strategy involves formulating a set of safety rules and deriving a control law that minimizes levels of unsafety, with the primary aim of reaching the desired destination even in the presence of violated safety rules [194, 221]. To consider uncertainties as a result of occluded areas, especially in intersections, RL can be used to learn safer policies according to a risk-based reward function [222]. Paying attention to the execution uncertainties of a certain trajectory, recent works address this topic in different ways. A framework akin to Linear-Quadratic Gaussian (LQG) is utilized to estimate uncertainty during the execution of a given candidate trajectory [223]. For other traffic participants, their control inputs are estimated using a local planner, and a state distribution is predicted using a Kalman filter [224].

2.3.2 Trajectory Prediction

Predicting the intentions and trajectories of road users has a significant impact on decision-making and trajectory planning [225]. The prediction task yields some inherent uncertainties due to the humans being involved that cannot be encoded into the AV software so far [226]. Therefore, numerous approaches to the topic of AV trajectory prediction address this challenge in the literature. Existing reviews categorize the related works into physics-based, maneuver-based, and interaction-aware prediction [227]. However, a categorization on the basis of the underlying method is also conceivable, which results in physics-based, pattern-based, and planning-based methods [228]. Following the second categorization, the remainder of this section will briefly present the associated approaches to trajectory prediction.

Physics-based

Physics-based models consider that the motion of road users depends only on the laws of physics. Therefore, future trajectories are predicted using dynamic and kinematic models [229]. Assumptions by using models like Constant Velocity (CV) [230] or Constant Turn Rate Acceleration (CTRA) are usually made here. To include uncertainty, Gaussian noise simulation can be used [231], which is commonly used to model the vehicle's state [167, 232]. A standard method for recursively estimating a vehicle's state from noisy measurements is the Kalman filter [224].

Dynamic and kinematic models are also the basis for predictions based on reachability analysis. In this approach, all states that a road user can physically reach are calculated and described in a so-called reachable set. These spatio-temporal sets are intended to be avoided by the AV. However, with larger prediction horizons, these sets become disproportionately extensive, which could lead to situations where no movement of the AV is possible anymore [233]. For this reason, next to the physical constraints, legal constraints based on traffic rules, for example, have been added [234, 235]. Further approaches, such as the RSS, assign corresponding responsibilities and intend to guarantee safety for a given trajectory [78]. However, finding reasonable parameters so that neither traffic flow is hindered nor collisions occur remains a challenging task here [236]. Initial criticism of this responsibility-based approach coming from the first AVs in the United States appears, stating that AVs may rather avoid blame than accidents [237]. To avoid the assumption of strict rule adherence of all road users, the initial approaches extend formal methods by considering the errors or traffic rule violations of other human road users [238]. Alternative approaches relieve these hard constraints by including stochastic information in so-called stochastic reachable sets [239].

Physics-based motion models are constrained in their capacity to predict motion for short durations, typically encompassing intervals of less than a second [227]. Notably, these models exhibit limitations in their ability to anticipate alterations in the road user's motion resulting from the execution of specific maneuvers or variations induced by external influences.

Pattern-based

Previous work has shown that pattern-based approaches outperform physics-based approaches with rising prediction horizons [240]. Leading algorithms of various prediction challenges, such as nuScenes [241] or Argoverse [242], underline this. These approaches are particularly relevant for the trajectory prediction of pedestrians [243]. The majority of these approaches comprise either Convolutional Neural Networks (CNNs) [244], Recurrent Neural Networks (RNNs) [245, 246] or a combination of both [247]. Other approaches represent the proximity between road agents using a dynamic weighted traffic graph [248]. All these approaches use the track history of a vehicle, process it in a neural network with encoder-decoder architecture, and predict the future trajectory of the vehicle. The primary distinctions among these approaches emerge from the utilization of diverse input data sources, often stemming from variations in scenario representations, as well as differences in their output predictions, which can include intentions, unimodal or multimodal trajectories [240]. To consider uncertainties, the trajectories as outputs are modeled using Gaussian distributions [249] or Dirichlet distributions [250]. An alternative approach by [251] pays particular attention to interactions and does not predict every vehicle separately. Instead, all vehicles within a road scenario are simultaneously considered at once and thereby account for interactions between different vehicles. Further approaches extend the input data to raw sensor data, such as Red-Green-Blue (RGB) images or lidar [252–254]. Generated Birds-Eye-View (BEV) images are also used to improve the scene understanding of the neural network [255, 256].

Planning-based

Behind the approaches to planning-based prediction is the idea of the implied similarity between the prediction task and the planning task. While algorithms for trajectory planning strive to determine a trajectory for an AV for a given state, the trajectory prediction strives to determine a trajectory for a third-party road user. One of the main differences is that in the case of prediction, the planning objective, as well as the goal state is unknown. Therefore, one major approach to planning-based prediction is to reason about the most likely goal and corresponding policy of an agent being predicted [228]. The concept of learning from demonstration involves inferring motion predictions by either estimating the underlying cost function or directly ascertaining the optimal policy. Therefore, learning the feature-based cost function as a planning objective for pedestrians

has been shown to generalize well [257]. A similar approach was extended to vehicle trajectory prediction using Inverse Reinforcement Learning (IRL) to learn the cost function [175, 258]. Improved prediction accuracy was obtained when the collision risk is explicitly considered [259]. Further methods involve using Behavior Cloning (BC) [260] and Generative Adversarial Imitational Learning (GAIL) [261, 262]. A common drawback of these approaches, which is a current field of research, is the large amount of data that is needed to train these algorithms [228].

2.4 Conclusion and Research Gap

In the state of the art, ethics and software development of AVs are mainly considered as separate topics, and a closer working relationship between ethics and software development is motivated [263]. On the one hand, there are ethical considerations that are technically difficult to implement or completely lack a technical dimension layer. Other approaches have shown to be of little relevance for practical use since they contradict the legislation. On the other hand, there are high-level requirements on the part of the legislation, which have not yet been implemented in the state of the art on the algorithmic side. The literature review has shown the particular importance of an interdisciplinary approach in advancing this endeavor [264].

Two main criteria became apparent as requirements to evaluate the current works in this field: First, ethical decision-making requires mature AV software for real-world application. Current works range from not considering technical aspects at all to existing prototype software. Instilling a computer program with the flexible intelligence necessary for ethical analysis is considered a challenging task [265]. The second criterion is the applicability of the proposed approach to ethical decision-making in terms of driving scenarios. Especially the discussion of the trolley problem has shown that solutions must not be limited to hazardous situations. Instead, any algorithmic decision of AVs must be considered from an ethical point of view [90]. Figure 2.6 categorizes the previously presented approaches from different ethical theories, namely deontology, consequentialism, virtue ethics, and risk ethics, according to these two dimensions. Only four works of the literature review provide prototype software of an ethical approach to AV decision-making. However, the proposed approaches here are tailored to specific scenarios. In general, the applicability of current approaches is also mainly limited to specific scenarios. There are only three works that follow a generalistic approach with applicability to mundane traffic situations. However, these works do not provide AV software. In particular, to the best of the author's knowledge, there is no work yet that covers both aspects to a sufficient degree. This opens a research gap for this thesis.

In addition to these two criteria, there is another crucial component that the existing literature has not considered. Legislation has recently imposed some relevant requirements that must be considered in ethical decision-making. In total, three sources were examined for the German and European scope, and twelve specific requirements were derived in Section 2.2.4. No work actively considering and meeting these requirements can be found in the state of the art so far.

The most advanced approaches in terms of software maturity and general applicability by Wang et al. [80] and Lindner et al. [84] both build up on deontology as ethical theory and implement rulesets. The requirements emerging from the German ethics committee, a European expert group, and the EU regulation suggest the use of risk ethics, e.g., by explicitly demanding to manage dilemmas by shared ethical principles for risk distribution (R6) [11]. Moreover, a review of the prior research in the field has led to the exploration of assessing risks within an ethical framework, which appears to be a promising approach.

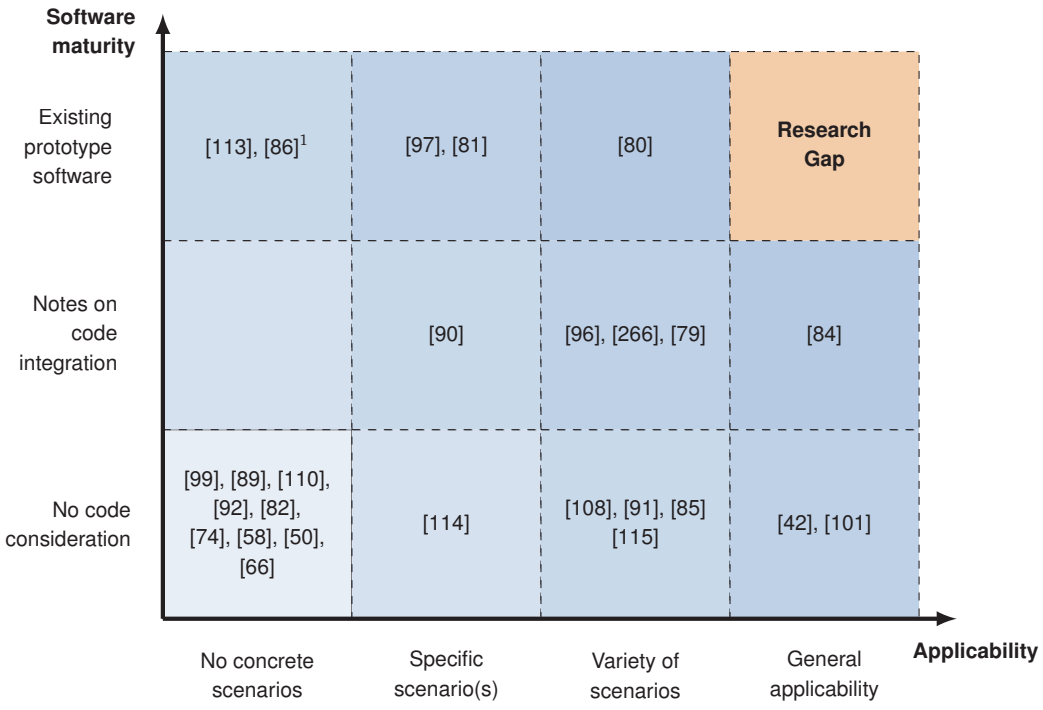


Figure 2.6: Systematic literature review on current approaches to ethical AV decision-making. Evaluation according to software maturity and applicability reveals the research gap for this thesis.
([1]: Application to an unmanned aircraft)

In summary, an AV software algorithm for ethical decision-making based on legislation with applicability to mundane traffic situations cannot be found in the current state of the art [267], which represents the research gap of this thesis. Risk ethics has been identified as a promising approach which has hardly been followed by the literature. To fill this gap, the following research question is posed to guide this thesis:

► *How can an ethical algorithm with fair risk assessment be realized for the trajectory planning of AVs?*

For a better structure, the superordinate research question results in five sub-questions, each of which addresses the individual aspects of the superordinate task:

Q1: How can ethical aspects be considered in algorithms for AV decision-making based on legal requirements? (addressed by Sections 3.1 and 3.2)

Q2: Which ethical principles may be considered in the trajectory planning of AVs? (addressed by Sections 3.3 and 3.4)

Q3: How can an ethical concept from several ethical principles be implemented in an algorithm for AV trajectory planning? (addressed by Chapter 3)

Q4: How can the aspect of fairness be taken into account in the decision-making process of AVs? (addressed by Chapter 4)

Q5: What are the actual effects of using an ethically motivated algorithm for trajectory planning compared to state of the art? (addressed by Chapter 5)

In the following, the research subject is examined on the basis of these research questions. Section 6.1 reviews at the end of this thesis to what extent these questions could be answered.

3 Development of a Motion Planning Framework with Risk Assessment

The preceding section has highlighted an evident research gap that warrants focused investigation. Consequently, this thesis focuses on the formulation of an ethical algorithm to govern AV trajectory planning. This algorithm, while multifaceted, primarily aims to harmonize technological advancement with ethical considerations. Therefore, a holistic approach is important to ensure that ethical principles guide the decision-making process. This endeavor encompasses an array of key requirements, each contributing to a comprehensive framework that not only addresses the complexity of AV decision-making but also takes ethics into account.

At the core of this initiative lies the need for an algorithm that reflects the complex nature of AV decision-making. Uncertainties inherent in real-world scenarios must be taken into account, allowing the algorithm to make informed decisions in the face of ambiguity. Beyond this, the algorithm must proficiently handle mundane traffic situations, seamlessly integrating into everyday scenarios instead of focusing only on critical scenarios. A primary consideration is that the development process and the algorithm itself adhere to European and German legal frameworks, as well as their corresponding recommendations. Next to the key requirements that guide the development of such decision algorithms, there are further aspects to consider. Transparency is another key requirement for ethical decision-making. The algorithm should be designed to make the decision-making process explainable and comprehensible to relevant stakeholders. This transparency extends to the adaptability of the algorithm. It should be able to flexibly account for various cultural contexts and individual preferences, ensuring that AV interact harmoniously with diverse societies. However, the ultimate validation of AVs hinges on societal acceptance. Beyond its technical foundation, the algorithm should resonate with the values and concerns of the community it serves.

In order to take the requirements regarding complexity into account in AV decision-making, the inclusion of uncertainties is essential, as shown in Chapter 2. Therefore, the application of risk ethics as an underlying ethical theory to AV decision-making appears to be a promising approach. To enable the application of risk ethics, technical foundations must be established first. For this purpose, in this chapter, a framework will be presented that focuses on the consideration of uncertainties to integrate ethical principles.

In the following, an algorithm for motion planning based on these requirements will be presented, and an answer will be given to the research questions Q1, Q2, and Q3. To structure this chapter, the following Section 3.1 will first give an overview of the proposed trajectory planning algorithm. The subsequent sections will then go into the details of each step and highlight the proposed approach to considering ethics. The consequent compliance with the requirements is discussed in Chapter 6.

The design of the algorithm was the core of the author's publication [268], and open-source software was made publicly available [269]. The aspect of maximum acceptable risk was previously covered more in detail by [270]. The proposed prediction model was part of the author's publication [271].

3.1 Motion Planning Framework with Risk Assessment

Section 2.3 presented various approaches to trajectory planning, such as graph-based, sampling-based, and optimization-based techniques. This work builds on a sampling-based approach since, in contrast to optimization-based algorithms, there are no restrictions on the formulation of the planning objective, and in contrast to graph-based algorithms, computation time goals can be achieved more easily for real vehicle applications. In addition, the sampling-based enables transparency due to the separation of trajectory generation and evaluation. The trajectory generation, though not the focus of this work, is inspired by the works of Werling et al. [200, 201, 272], who generated jerk-optimal trajectories using a Frenet coordinate system.

The developed algorithm consists of four steps, as shown in Figure 3.1:

1. Multiple trajectories are sampled according to a pre-specified discretization scheme, given the current state of the AV. Various approaches for generating the trajectories are feasible here. In this work, the trajectories are generated in a Frenet coordinate system [272]. Furthermore, the algorithm assigns a risk value to each road user for each trajectory, which allows consideration of risk distributions in trajectory planning as a key aspect of ethical decision-making. (See Section 3.2)
2. The sampled trajectories are classified into four distinct validity levels. The primary objective is to prioritize trajectories at the highest validity level, thereby guaranteeing existing solutions, even if not all quality criteria are satisfied. From an ethical point of view, the deliberation of a risk threshold is an active consideration here, referred to as the maximum acceptable risk. (See Section 3.3)
3. For those trajectories within the highest available validity level (for example, ‘valid’), an ethical cost function is calculated. The algorithm contains different cost functions depending on the validity level. Among other critical dimensions, the cost function aims to ensure a fair distribution of risk. Various ethical aspects that are taken into account will be illuminated. (See Section 3.4)
4. The trajectory from the highest available validity level and the lowest calculated cost is selected and executed. After a single timestep is completed, a new trajectory is planned for the next timestep starting from step 1.

These four steps provide the framework for the algorithm and the remainder of this chapter. Each following section will elaborate on these steps in detail.

3.2 Trajectory Sampling & Risk Quantification

In order to sample the trajectories in the first step of the trajectory planning algorithm, a reference path is required, which fulfills the navigation task in the three-layer model of the primary driving task [273]. This reference path consists of a sequence of waypoints and represents a connection between an initial and a target position. It does not include a velocity profile, does not take objects into account, and is, accordingly, not collision-free. In the algorithm, the global navigation is performed by a hierarchical search along the centerlines in a lanelet network. Further details regarding the reference path generation can be found in [274].

Given the reference path, the first step of the sampling-based trajectory planning approach is to generate multiple trajectories from a given state of the AV. The approach followed here on trajectory sampling builds up

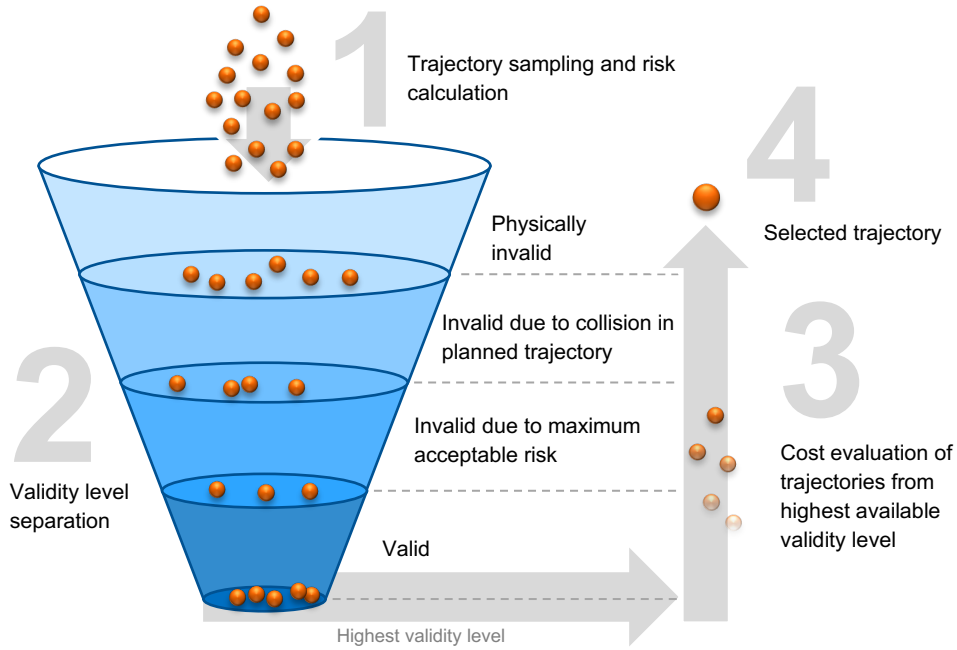


Figure 3.1: Overview of the proposed ethical trajectory planning algorithm in four steps. The small orange balls symbolize trajectories that are sampled in the first step. Next, the trajectories are subjected to validity checks like in a filter screen visualized here (Step 2). Only those trajectories of the highest available validity level (here: five trajectories from 'valid') are assigned costs, whereas higher costs are represented with higher transparency (Step 3). In the last step, the trajectory with the lowest cost is selected. The figure was adapted from a previous author's publication [268].

on previous works [200, 201, 272]. Hence, the trajectory generation is subdivided into longitudinal and lateral movement, using a Frenet coordinate system: instead of using the cartesian coordinates (x, y), any location can be described unambiguously with an arc length s and a lateral displacement d along the reference path. To generate jerk-optimal trajectories, fifth-order polynomes are used to describe the lateral movement and fourth-order polynomes for the longitudinal movement [201].

First, to sample various options for lateral movement, the goal state at the planning horizon varies along the d dimension. Figure 3.2 a) shows the lateral movement with discretization for five exemplary values of d. Second, to sample various options for longitudinal movement, the target velocity at the planning horizon is varied. Various goal velocities can be generated as linear interpolation using the minimum and maximum reachable velocities given the initial state. To allow smooth velocity keeping, longitudinal movement without acceleration is added, where the goal state's velocity equals the initial velocity. This results in the longitudinal profiles as shown in Figure 3.2b).

Finally, the combination of lateral and longitudinal movements results in a predefined number of trajectories as in c) due to the trajectory sampling step. In addition, the planning horizon can also be varied to generate further variations of trajectories. However, evaluating trajectories with different planning horizons leads to difficulties in terms of comparability, which is why a fixed planning horizon of $2.0\,s$ is used here. Preliminary experiments to investigate the planning horizon have shown that $2.0\,s$ is a reasonable compromise between computation time and planning performance. Longer horizons of more than $2.0\,s$ did not show significant improvements but would lead to more computation time.

The next step is to assign a measure for a related risk to each sampled trajectory. There are various definitions of risk available [275]. This thesis defines risk as an expected value consisting of two key components:

- *Probability*: Risk involves the consideration of the likelihood of an event happening. This can be expressed as a percentage, a fraction, or any other relevant measure of probability.
- *Consequences*: Risk also takes into account the potential outcomes or consequences of the event. This includes evaluating the severity and extent of possible negative impacts.

Accordingly, risk is defined as the product of a probability that a certain event will occur and a measure of the consequences of that event. This is particularly important since previous work showed that for an ethical assessment, both dimensions have to be taken into account [92]. In the case of autonomous driving, risk R can therefore be defined as the product of a collision probability $p_{\text{collision}}$ and the estimated harm out of that potential collision H as in Equation 3.1 given a trajectory u at timestep t. The calculation of collision probabilities resulting from prediction uncertainties is described in detail in Section 3.2.1, while Section 3.2.2 addresses the harm estimation model.

$$R(u,t) = p_{\text{collision}}(u,t)\,H(u,t) \qquad p_{\text{collision}}, H \in [0,1] \tag{3.1}$$

Given the harm and collision probability at each state of a planned trajectory results in a time-variant risk along the planning horizon for each sampled trajectory. However, the risk being used as an unambiguous selection criterion requires the characterization of every possible trajectory by a single risk value. This becomes challenging because the collision probabilities along the planning horizon, and thus, the risks for a collision of the same road users are not independent. Therefore, assuming purely dependent potential collisions the risk $R_{\text{Traj, dep.}}$ of a trajectory u is denoted as the maximum risk over time:

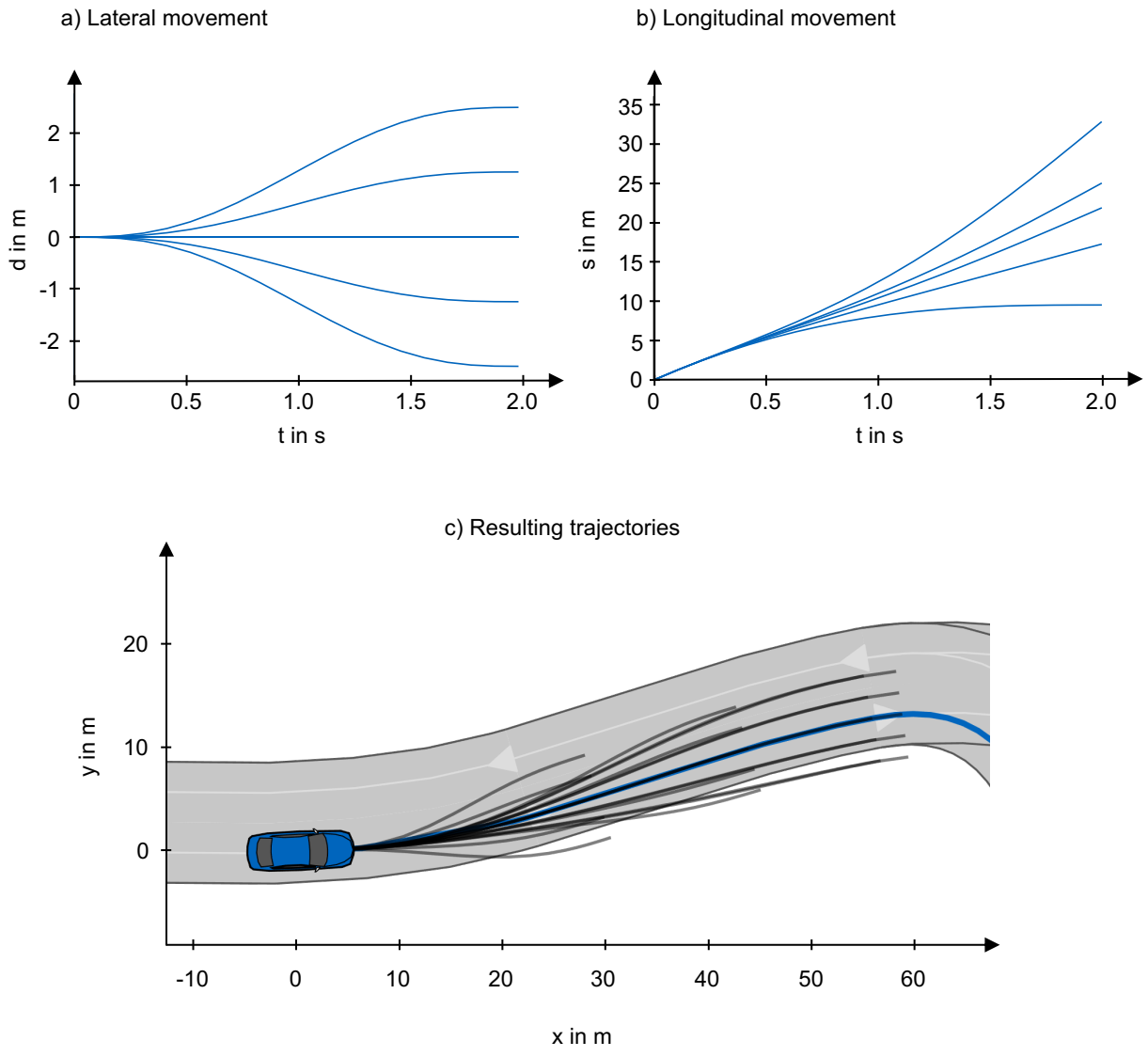


Figure 3.2: The sampling of trajectories in a Frenet coordinate system is separated into a) the lateral movement and b) the longitudinal movement. The combination of those variations results in the trajectories in c).

$$R_{\text{Traj, dep.}}(u) = \max_{t \in [t_0, t_{\text{end}}]} (R(u,t)) \tag{3.2}$$

Using the maximum value of risk represents an approximation that neglects the effects of independent risks originating from a single road user. However, since collision probabilities originate from a neural network model, as will be shown in Section 3.2.1, this approximation cannot be resolved.

Risks originating from potential collisions with different road users are assumed to be independent and therefore calculated with Equation 3.3, where S_R denotes a set of risks containing a calculated risk value for each road user. Equation 3.3 assumes that the independent risks are calculated analogously to their underlying collision probabilities. To calculate the probability for (at least) one collision given several independent collision probabilities, the probability for no collision is therefore calculated first as a product of probabilities. The probability of a collision is then the counter-probability of this product. Here, as well, assuming independent risks neglects the aspect of road user interactions.

$$R_{\text{Traj, indep.}} = 1 - \prod_{|S_R|} \left(1 - R_{\text{Traj, dep.}}\right) \tag{3.3}$$

This allows the assignment of a single risk value for every road user for each generated trajectory as schematically visualized by Figure 3.3. The presented risk calculation creates a link between a candidate trajectory (A, B or C in Figure 3.3) and a resulting risk distribution among the road users (R_{ego}, R_1 and R_2 in Figure 3.3). Thus, the theoretical problem of risk distribution is established in the AV application and forms the basis for further ethical considerations.

3.2.1 Collision Probability

The probability of potential collisions in autonomous driving arises from a variety of uncertainties. These uncertainties encompass aspects of environmental perception, such as the detection and localization of other road users, as well as uncertainties related to vehicle control. The primary emphasis of this study centers on addressing one of the most pivotal sources of uncertainty in trajectory planning, which is the prediction of trajectories for other road users. In addition to uncertainties of the underlying model, the prediction task also consists of inherent uncertainties. Even the best possible prediction model can not predict the movement of other road users with any degree of accuracy because the (human) behavior of other road users is stochastic [228]. From a model perspective, this means that there are multiple possible outputs for a single observation as input representation.

There are various methods to quantify the uncertainties that are associated with the trajectory prediction of other road users. A neural network serves as a pattern-based method for the prediction task, which the author published in a previous publication with open-source software called *Wale-Net* [271, 276]. Figure 3.4 shows the architecture of the prediction network that builds up on the idea of Convolutional Social Pooling [247] and was adapted for the use in structured road networks.

The previous positions (up to $2\,s$) of the predicted vehicle are encoded using a Long-Short-Term-Memory (LSTM) layer. The same layer also encodes the past positions of third-party road users around the predicted vehicle that are arranged in a grid representation. Using this encoded tensor enables the representation of first-order interactions between road users. In addition to the dynamics of relevant road users, the network model pays attention to the road network geometry. Therefore, a convolutional layer generates and processes a pixel image of a lanelet representation [277]. Following the approach of encoder-decoder

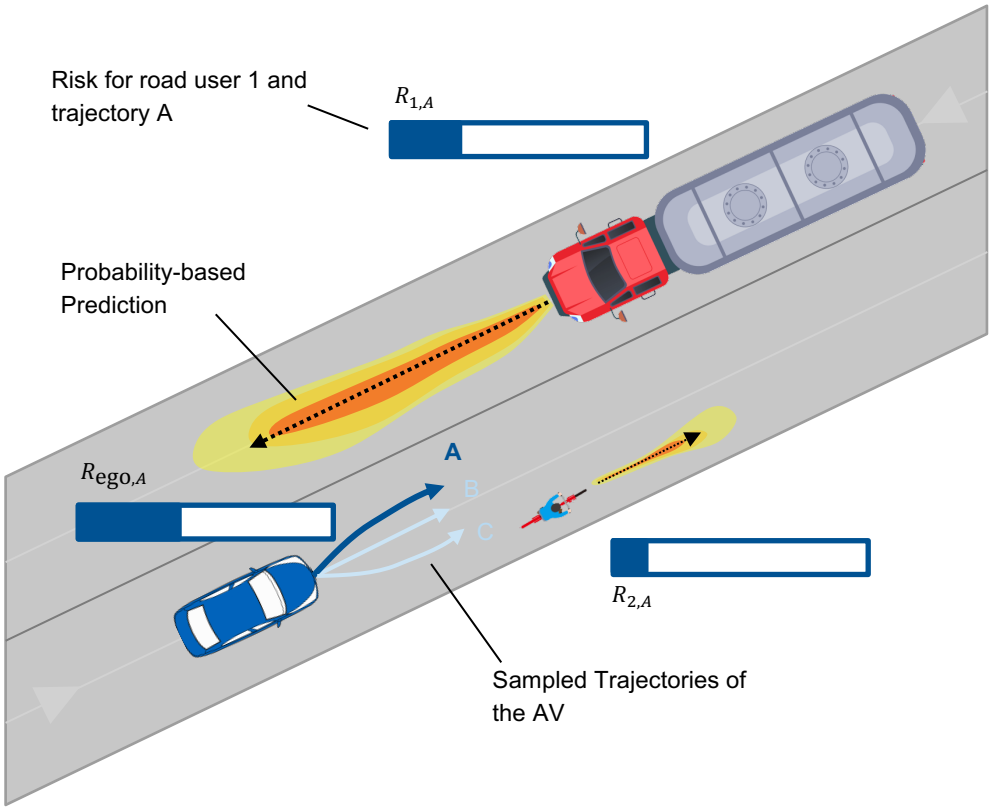


Figure 3.3: Schematic visualization of trajectory planning based on risk distribution. On the basis of a probability-based prediction of all road users (shown here as heatmaps around the black most likely predictions) and an estimated harm value, every trajectory of the AV can be assigned risk values for every road user. The figure was adapted from a previous author's publication [268].

neural networks [278], the various encoded information is concatenated in the latent space. Trajectory predictions are generated using subsequent LSTM and fully-connected layers. The output representation consists of a variant number of timesteps according to the prediction horizon with a fixed step size of $0.1\ s$. It contains the expected position values $\mu_x(t)$ and $\mu_y(t)$ at the prediction timestep t together with a probability representation as a bivariate normal distribution. Therefore, at every timestep together with an expected position, a $2x2$ covariance matrix $\Sigma(t)$ is predicted with the standard deviation in x and y, denoted as $\sigma_x(t), \sigma_y(t)$, and the correlation coefficient $\rho(t)$.

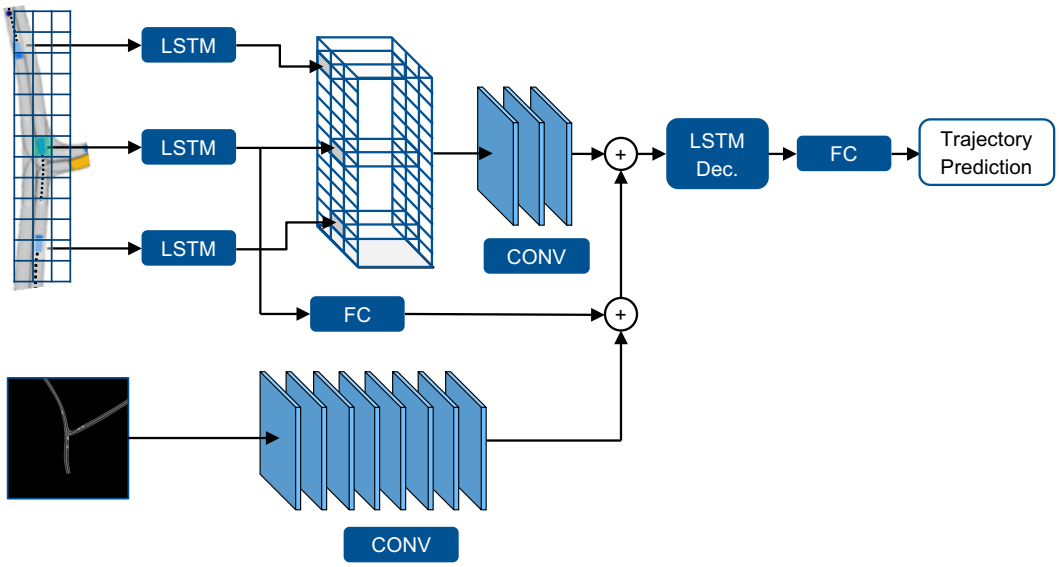


Figure 3.4: Schematic overview of the architecture of the probabilistic prediction model Wale-Net [271]. Taking a map as a pixel image and the past trajectories of relevant road users in a grid as input, the neural network model predicts the future trajectories with uncertainties as an output.

In order to represent the uncertainties in the training process, a Negative Log Likelihood (NLL) is used as a loss function, described by Equations 3.4 and 3.5, where the ground truth positions are described by x_{GT} and y_{GT}. The NLL is calculated as an average over the prediction horizon of n timesteps.

$$\mathrm{NLL} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{z_i}{1-\rho_i^2} - \ln\left(\frac{1}{\sigma_{x,i}\,\sigma_{y,i}\,\sqrt{1-\rho_i^2}}\right)\right] \tag{3.4}$$

$$z = \frac{(x_{GT}-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x_{GT}-\mu_x)(y_{GT}-\mu_y)}{\sigma_x\,\sigma_y} + \frac{(y_{GT}-\mu_y)^2}{\sigma_y^2} \tag{3.5}$$

The network was adapted and trained with various input lengths to avoid requiring a long observation period of, for example, $2\ s$, as in [247]. Consequently, a probability-based prediction can be made from the first observation step on.

The uncertainty representation as a bivariate normal distribution of the trajectory prediction builds the basis for the collision probabilities. The Probability Density Function (PDF) based on the covariance matrix as prediction output is described by Equation 3.6 and visualized in Figure 3.5. In this case, the ground truth variables x_{GT} and y_{GT} to determine z in Equation 3.5 must be replaced by the distribution variables X and Y.

$$PDF(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{z(X,Y)}{2(1-\rho^2)}\right] \tag{3.6}$$

To consider the shape of the prediction object, the PDF is modified. Therefore, it is composed of three separate distributions: In addition to the original PDF that has its expected values μ_x and μ_y at the center of gravity of the object, two additional partial distributions are employed to represent the front and rear of an object. The collision probability can then be calculated given the ego vehicle's shape based on the three PDFs [274]. The resulting PDF adapted to the vehicle's shape and an exemplary ego-vehicle is shown by Figure 3.5. The collision probability can then be calculated given the ego-vehicle's shape described by x_1, x_2, y_1, y_2, and the modified PDF according to Equation 3.7. This method is based on the assumption that a state is a collision state if and only if the shapes of the two considered objects are intersected. Theoretically, however, there are further states of the predicted object that would not be reachable without a collision with the planned state of the ego-AV. An exact calculation of the collision states based on physical constraints is not performed here due to the requirement for low computation time.

$$p_{\text{collision}} \approx p(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} PDF(X,Y)\,\mathrm{d}x\,\mathrm{d}y \tag{3.7}$$

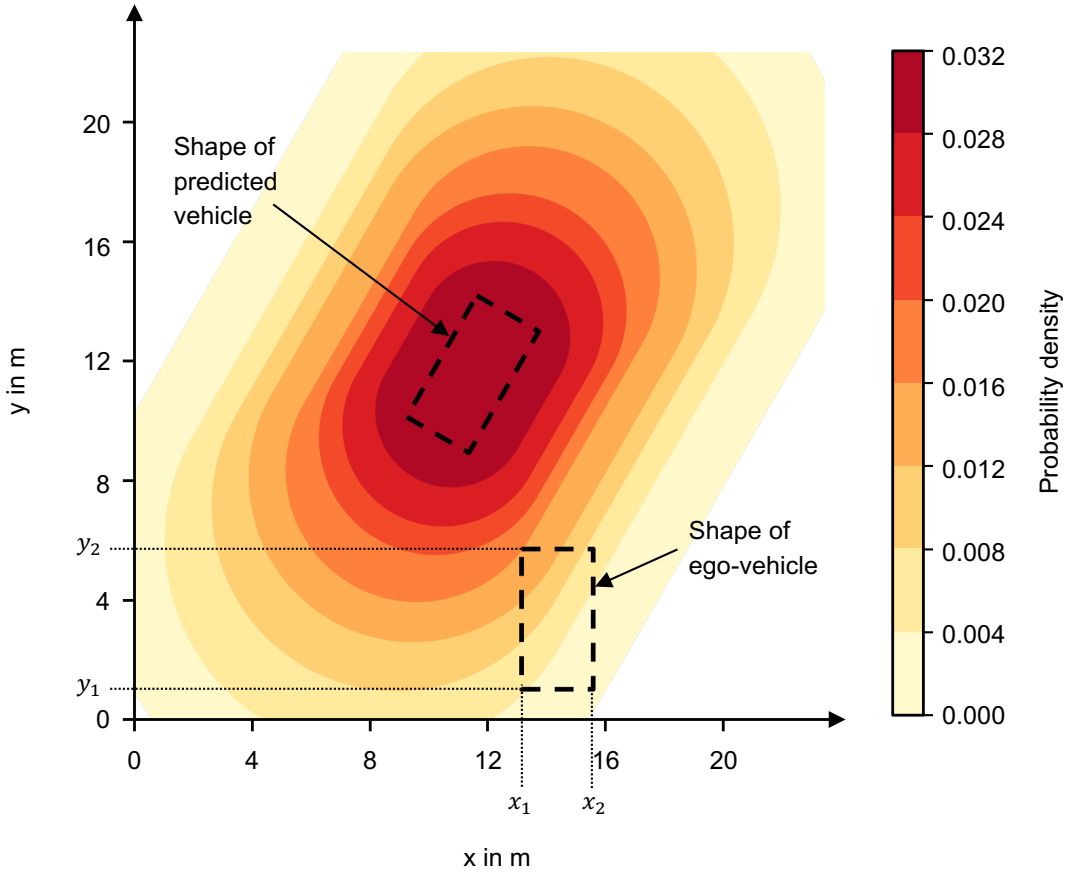


Figure 3.5: Calculation of collision probabilities resulting from bivariate Gaussian uncertainties and the shapes of the object. The figure was modified according to [274].

3.2.2 Harm Estimation

Estimating the harm of a potential collision is not only a technical challenge but also has ethical implications. Quantifying and weighing different types of harm against each other presents enormous ethical difficulties, particularly in the case of extreme accidents that could be fatal. The ethical guidelines and legislation, as discussed in Section 2.2.4, provide the basis for the harm model: Firstly, the proposed harm model only refers to personal injuries and does not consider property damage, as the legislation mandates a clear prioritization here (R2). Secondly, the model must not perform any assessment based on discriminatory factors, such as gender or age (R4 and R10). Finally, the vulnerability of various road users must be considered (R5 and R12), which suggests distinguishing between protected and unprotected road users.

To make ethical considerations when quantifying harm, precise knowledge of accident consequences seems necessary as a technical requirement. However, accurately predicting the severity of an accident in practice is limited. AVs have only limited information to determine the severity of an accident. For example, the number and location of occupants in a vehicle and vehicle-specific safety elements are unknown. Hence, the modeling of accident severity can only be based on known characteristics with physical relationships. As the harm model will evaluate vast amounts of possible collisions at each trajectory planning step, the calculation time is important to consider here, which restricts the available model complexity.

The proposed harm estimation model aims to estimate the personal damage of a potential collision and map it to a scale of values from 0 to 1. 0 denotes a collision without any harm (to humans), and 1 accounts for the greatest possible harm. Therefore, any subjective factors are excluded in the modeling, such as quality of life [279], and only objective factors are considered following the approach of a kinetic energy model. The severity of injury increases in proportion to the kinetic energy. Relevant studies show that kinetic energy serves as a robust indicator of harm [280]: In general, when a road user experiences a higher level of kinetic energy during an accident, the resulting injuries tend to be more severe. Similarly, the probability of death in an accident rises with higher velocities and thus higher kinetic energies [281]. The kinetic energy that a road user encounters in the event of a collision depends on five primary factors: velocities and masses of the two colliding parties, impact angle, and impact area, and whether the road user has any protection (e.g., in the case of a car) or not (e.g., pedestrians).

To map these input variables to a measure for estimated harm, the accident database of the National Highway Traffic Safety Administration's Crash Report Sampling System (NHTSA) [282] provides relevant crash data from the United States. In addition to the discussed input variables, it provides the severity of an accident. The severity of an accident is described using the Abbreviated Injury Scale [283]. In particular, the probability $P_{\mathrm{MAIS3+}}$ that an accident leads to injuries with the severity of Maximum Abbreviated Injury Scale (MAIS) of three or greater expresses the severity of an accident on a scale of 0 to 1 as a harm metric. The underlying assumption here is that the probability MAIS3+ correlates with the severity of a potential crash. The physical relationships depicted in the logistic regression support this assumption.

In line with the recommendations of the German Ethics code and the EU legislation, the proposed model distinguishes between protected (vehicles, trucks, etc.) and unprotected (pedestrians, cyclists, etc.) road users. Logistic regression for both cases determines the empiric parameters of a logistic regression model, described by Equations 3.8 and 3.9. m and v are the mass and velocity of the two road users A and B, α is the collision angle, and c_0, c_1 and c_{area} are the empirically determined coefficients. Further details, especially on estimating the vehicle's mass, impact angle, and area of a potential collision, along with the regression method and variations in modeling complexity, can be found in [284].

$$\Delta v_A = \frac{m_B}{m_A + m_B} \sqrt{v_A^2 + v_B^2 - 2\, v_A v_B \cos\alpha} \tag{3.8}$$

$$H = P_{\text{MAIS3+}} = \frac{1}{1 + e^{c_0 - c_1 \Delta v - c_{\text{area}}}} \tag{3.9}$$

Figure 3.6 shows the $P_{\text{MAIS3+}}$ over the velocity change Δv during a crash considering the newtonian mechanics from Equation 3.8. As can be seen, in the case of no protection, the same values for Δv lead to higher values in harm. For protected road users, the impact area has an influence on the accident severity. However, the database is biased toward a driver: performing the logistic regression with various impact areas leads to the result that crashing a car on the driver's side is worse than the same crash on the other side. This is due to the fact that statistically, the driver's seat is occupied more often.

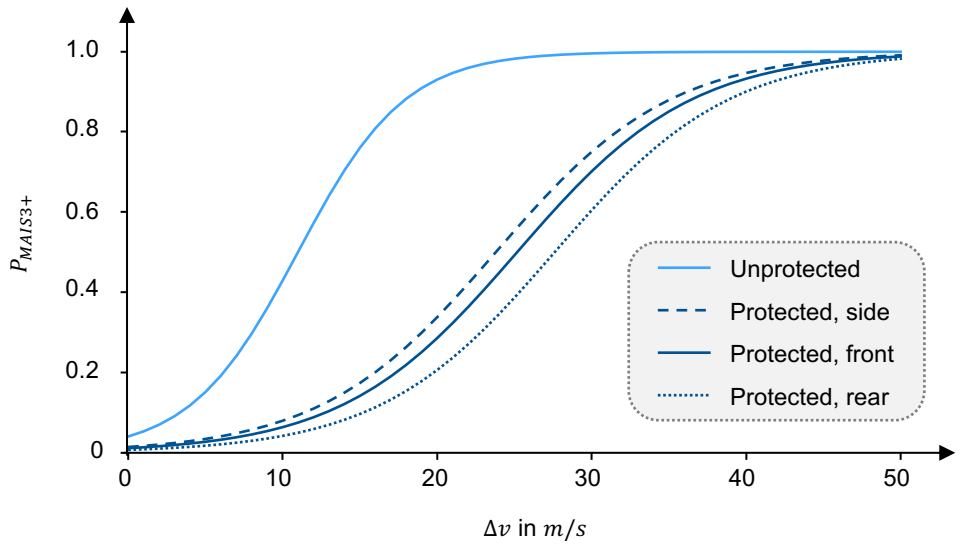


Figure 3.6: Course and sensitivity of the harm model as logistic regression over Δv for various impact areas and the unprotected case based on the data provided by [284].

Aligning the impact areas symmetrically around the horizontal axis, as shown in Figure 3.7, mitigates this bias. A disadvantage of the logistic regression model is the inaccuracy near $\Delta v = 0$. Here, the expected harm must be 0 from a physical point of view, which is not mapped due to the asymptotic property of the function. Because no collisions are expected at $\Delta v = 0$, since both objects are then at a standstill, this circumstance is not expected to be of significance.

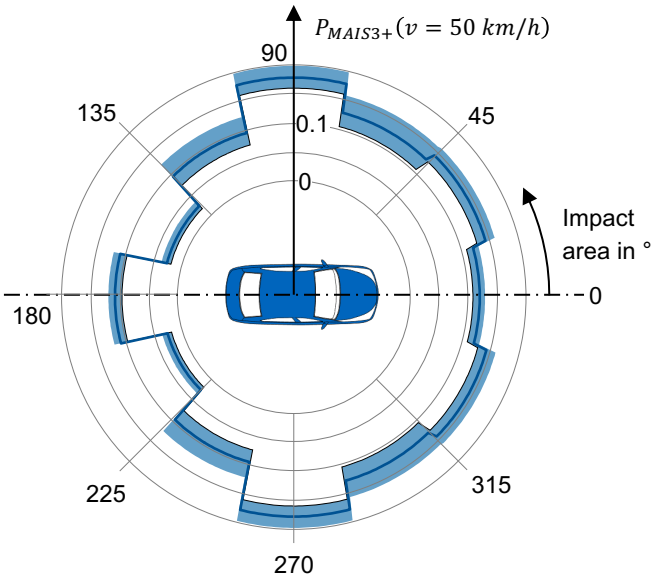


Figure 3.7: Harm values for $\Delta v = 50 km/h$ with varying impact angles. The blue areas indicate the 95% confidence interval based on the data provided by [284].

3.3 Maximum Acceptable Risk

The previous sections established a framework for risk assessment in trajectory planning by creating a connection between a planned trajectory and the associated risks for related road users, which equals the first step in the overview of Section 3.1. According to that, the next step is to perform validity checks on each of the trajectories. While this is a well-known method in the literature, a novel concept of maximum acceptable risk is introduced as a validity check.

Before addressing the question of fair risk distribution, the question arises of how much risk should be considered acceptable in road traffic. The concept of setting thresholds to risks is well established and goes hand in hand with questions of the required safety of AVs, which is commonly discussed in the literature. However, the focus is to integrate such thresholds into the trajectory planning to guide the AV behavior.

The AV behavior, which is determined by the trajectory planning, thus has an impact on how much risk arises in a situation: whether the AV performs an overtaking maneuver or not or the velocity the AV chooses has a massive impact on the amount of risk that occurs for all related road users. For a method of answering this question of accepted risk using the human driver as a benchmark together with acceptance studies, the reader is referred to the author's publication [270]. However, the focus here will be on implementing such a concept of accepted risk in trajectory planning. The ultimate goal of this approach is to use risk as a criterion in AV decision-making and guide AV behavior in accordance with acceptable risk.

Therefore, the next step of the planning algorithm uses the sampled trajectories along with the calculated risk from step (1) and checks these trajectories for validity. Next to validity checks that are well known in the

literature, such as the physical feasibility of the trajectory or collision checks, the maximum acceptable risk serves as a further validity criterion in the second step of the trajectory planning algorithm. The set of all valid trajectories according to the accepted risk criterion is calculated as follows: Let U be the set of all trajectories generated by the trajectory planner in the sampling step (1) that are to be checked for validity (2). The goal is to determine the set U_V, which comprises all trajectories from which a final trajectory should be chosen (3), for example, by using a cost function. By applying the maximum acceptable risk R_{max}, the set U_V is obtained by Equation 3.10. If there is a set of trajectories that fulfills the criterion of maximum acceptable risk, the trajectory being chosen for execution must be of this set. Otherwise, if there is no trajectory that meets this requirement, the cost function evaluates the set of all trajectories.

$$U_V = \begin{cases} \{u \in U \mid R_u \leq R_{\text{max}}\} & \exists\, u \in U : R_u \leq R_{\text{max}} \\ U & \text{otherwise} \end{cases} \tag{3.10}$$

Given the fact that AVs will be involved in accidents, as shown by Chapter 1, it becomes evident that the AV will be in states where no trajectory can be found that fulfills the acceptable risk criterion. If no trajectory fulfills the condition of maximum accepted risk, the vehicle is in a high-risk situation. Usually, the trajectories are evaluated with various cost terms that are used in the literature [159] that can be related to either safety goals (J_{Risk}), comfort goals (J_{Comfort}), or mobility/efficiency (J_{Mobility}) goals. In the case of a high-risk situation, where a trajectory has to be chosen that does not meet the maximum accepted risk requirement, the cost function is evaluated with different parameters than under normal conditions. Cost terms with the objective of vehicle safety must be weighted significantly higher than those for comfort or mobility. The proposed implementation neglects all costs other than risk if the maximum acceptable risk is exceeded (Equation 3.11). The cost terms that are used in J_{Risk} will be explained in more detail in Section 3.4.

$$J_{\text{total}} = \begin{cases} w_R J_{\text{Risk}} & \\ \quad + w_M J_{\text{Mobility}} & \\ \quad + w_C J_{\text{Comfort}} & \exists\, u \in U : R_u \leq R_{\text{max}} \\ \\ J_{\text{Risk}} & \text{otherwise} \end{cases} \tag{3.11}$$

This method aims to provide guidance for AV decision-making. For example, as shown by Figure 3.8, it should stop the AV from performing an overtaking maneuver if it exceeds the accepted risk. The results section (5.3) will investigate this concept within trajectory planning.

3.4 Risk Distribution

The main goal of the development of an ethical trajectory planning algorithm is a fair risk distribution in contrast to a selfish risk distribution, as motivated in the introduction of this thesis. The question of what constitutes fairness or a fair risk distribution is not trivial and still concerns current research in the field of ethics. A well-established approach to fairness and distributive justice comes from the philosopher John Rawls (1921 - 2002) [20]. According to a social contract, he suggests that ignorance of one individual's circumstances enables more objective consideration of how societies should operate. This can be transferred to the application of risk distribution in autonomous driving: One key to a fair risk distribution could be asking

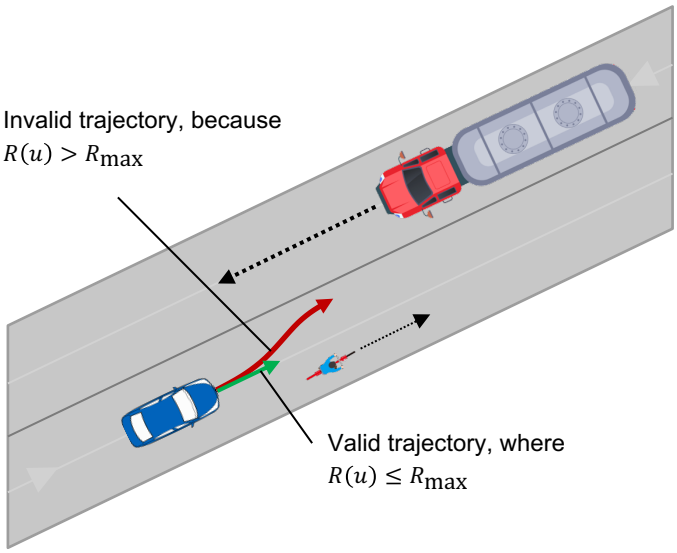


Figure 3.8: Exemplary scenario to showcase the concept of maximum acceptable risk, which is supposed to guide the behavior in trajectory planning. If the maximum acceptable risk is exceeded by the risk of a trajectory $R(u)$, a maneuver, such as an overtaking maneuver (red trajectory) here, should not be performed, but the AV should stay behind (green trajectory). The figure was adapted from a previous author's publication [268].

a representative group of people about what they consider a fair decision in terms of risk distribution without being personally involved in that fictive situation. This, however, requires a framework that creates a link between trajectory planning and such a survey. To break down the problem of decision-making, first, a model with predefined ethical principles is defined, of which the parameters can be set by means of such a survey. This top-down approach with variable parameters ensures explainability and transparency on the hand. On the other hand, it enables adaptability to different cultures, for example, as motivated at the beginning of this chapter.

Various distribution problems in the literature can be transferred to the application of AV decision-making. The distribution of monetary values [285], for example, or the allocation of scarce medical resources, such as in the case of organ donation [111] or during the Corona Virus Disease (COVID)-19 pandemic [286, 287] can serve as an example here. However, there are two important differences to consider: (1) While in most distribution problems, the amount of the distributive good stays constant (e.g., amount of money, organs, etc.), this is not the case in AV risk assessment. Different trajectory choices can lead to different total amounts of risk. (2) The distribution of risk is two-dimensional as risk includes a probability of occurrence and estimated harm (Equation 3.2). This two-dimensional character must also be considered.

Hence, an examination of diverse distribution problems from existing literature is conducted, and principles of distribution that align with an ethical theory are extracted. These principles are subsequently analyzed under two primary aspects. First, the principle should align with the legislation and ethical guidelines presented in Section 2.2.4. Second, the ethical principle must be feasible to implement in a trajectory planning algorithm, which requires a mathematical formulation.

From these points of view, four ethical principles emerge that can be taken into account for a fair distribution of risk. Recommendation R6, as presented by Section 2.2.4 suggests the use of shared ethical principles to risk distribution. The state of the art in trajectory planning shows that a linear cost function consisting of weighted cost terms, according to Equation 2.1 (Section 2.3), is a well-established method to combine several planning objectives. Therefore, the following four principles are integrated accordingly as a weighted sum in the cost function as part of the trajectory planning: the *Bayes principle* with costs J_B (Section 3.4.1), the *equality principle* with costs J_E (Section 3.4.2), the *maximin principle* with costs J_M (Section 3.4.3), and

the *responsibility principle* with costs J_R (Section 3.4.4). The following sections will describe these principles and their moral justification for using them to guide risk distribution. The according weighting parameters w_B, w_E, and w_M will be addressed in Chapter 4. A discussion regarding the selection of these principles will be provided in Section 6.2.2.

$$J_{\text{Risk}}(u) = w_B\, J_B(u) + w_E\, J_E(u) + w_M\, J_M + w_R\, J_R(u) \tag{3.12}$$

3.4.1 Bayes Principle

The Bayes principle from risk ethics demands maximization of the overall social benefit and conforms to a utilitarian demand [288, 289]. This corresponds to a minimization of the overall risk. The risk assessed to one person can be outweighed by the benefit done to another, which is explicitly in line with the German ethics committee [98]. According to previous studies, this principle is favored by most road users [7]. In the case of trajectory planning, the principle demands choosing a trajectory that minimizes the total risk of all road users according to Equation 3.13. Therefore, the corresponding costs J_B include the sum of the assigned risks of all road users. The total risk of all road users is then normalized by the number of road users that are considered in a specific scenario to assess the risk independently from the number of road users considered $|S_R|$. S_R describes the set of risks for each road user for a given trajectory u.

$$J_B(u) = \frac{\sum_{i=1}^{|S_R|} R_i(u)}{|S_R|} \tag{3.13}$$

However, using the Bayes principle as the ultimate decision criterion could lead to decisions that could be considered rather unintuitive. Figure 3.9a shows an example where the Bayes principle would choose an option with high risk to one person and zero risk to another (Option A). There are reasons to argue for choosing Option B with a more equal distribution of risk between the two involved and lower harms. Therefore, the European expert group advocated not only to use a single ethical principle but shared ethical principles [11].

3.4.2 Equality Principle

Furthermore, the equality principle is introduced to incorporate the concept of equal distribution. It is widely endorsed as the default principle in social settings where cooperation and harmony are primary goals [290]. This principle advocates for equality in risk distribution by minimizing disparities among the risks considered. It intends to prevent the Bayes principle from introducing a bias that disproportionately increases the risks for certain individual road users at the expense of reducing the overall risk. Combined with the harm model (Section 3.2.2), which specifically considers the vulnerability of VRUs, the equality principle addresses the EU expert group's requirement for redressing vulnerability inequalities among road users (R5 in Section 2.2.4) [11]. The costs according to the equality principle J_E are denoted by Equation 3.14. In the equation, the difference of all possible combinations of risk pairs is calculated and summed up as a measure of inequality. Other methods, such as the difference between the highest and lowest risk, would also be conceivable here. Similar to the Bayes principle, the costs are normalized by the number of added terms so that the costs range from 0 to 1 regardless of the number of road users.

$$J_E(u) = \frac{\sum_{i=1}^{|S_R|} \sum_{j=i}^{|S_R|} |R_i(u) - R_j(u)|}{\sum_{k}^{|S_R|-1} k} \tag{3.14}$$

Using only the equality principle as a single decision criterion could also lead to unwanted decisions. As Figure 3.9b shows, the equality principle would choose an option with two equal but extensively high risks over a small difference with very low risks. This could lead to decisions where the risks for both persons in a fictive example are higher than in the alternative option, which can be rejected from a rational point of view.

3.4.3 Maximin Principle

The Bayes and equality principles introduced so far only consider risk as a whole but neglect the two dimensions of risk: collision probability and harm. According to the risk definition given by Equation 3.3, the same risk value can result from a low probability for high harm and a high probability for low harm. The proposed framework, however, should allow separate consideration of collision probability and harm. The maximin principle enables decoupled assessment and goes back to Rawls' theory of justice [20]. It is considered to be the most effective solution to the problem of cooperation [291]. The maximin principle requires choosing the option with the lowest possible harm. When applied in the context of AVs, this principle entails the imperative of minimizing the most substantial harm, regardless of its probability. To achieve this, the principle demands selecting a trajectory where the greatest possible harm is minimal. This amounts to the idea of prioritizing the worst-off. The costs according to the maximin principle are given by Equation 3.15 and only consider harm without the according probability. Analogous to S_R, S_H describes a set of associated harms to road users for a given trajectory u. The costs for the maximin principle do not contain the usually low probabilities. Therefore, γ is an empirically determined discount factor that puts the average maximin costs in a similar range of values.

$$J_M(u) = \left[\max_{H_i(u)} (S_H)\right] \gamma \tag{3.15}$$

However, criticism from the literature remarks that the maximin principle gives undue weight to the moral claims of the worst-off [92]. Neglecting the probability of a potential outcome could lead to decisions in which very unlikely events of high estimated harm are avoided in favor of slightly less harm but much higher probabilities, as depicted by Figure 3.9c.

3.4.4 Responsibility Principle

Responsibility is a guiding principle for various approaches in the literature of AVs, which is argued that it cannot be ignored [58]. Responsibility of road users can arise from different reasons and is usually depicted in the form of rules. These rules can be based on physical limitations [292], common sense [78], or traffic laws [235]. They are intended to absolve AVs of any blame in the event of an accident. However, these approaches often neglect the inherent risks of driving in stochastic traffic conditions with other road users. First investigations show that relying solely on responsibility-based approaches could result in AVs prioritizing avoiding blame over preventing accidents [237]. This motivates the investigation of the responsibility principle as an additional ethical principle for risk management.

From an ethical standpoint, there are compelling reasons to incorporate responsibility-based approaches into a risk distribution framework. The risks in road traffic stem from the characteristics and actions of road users.

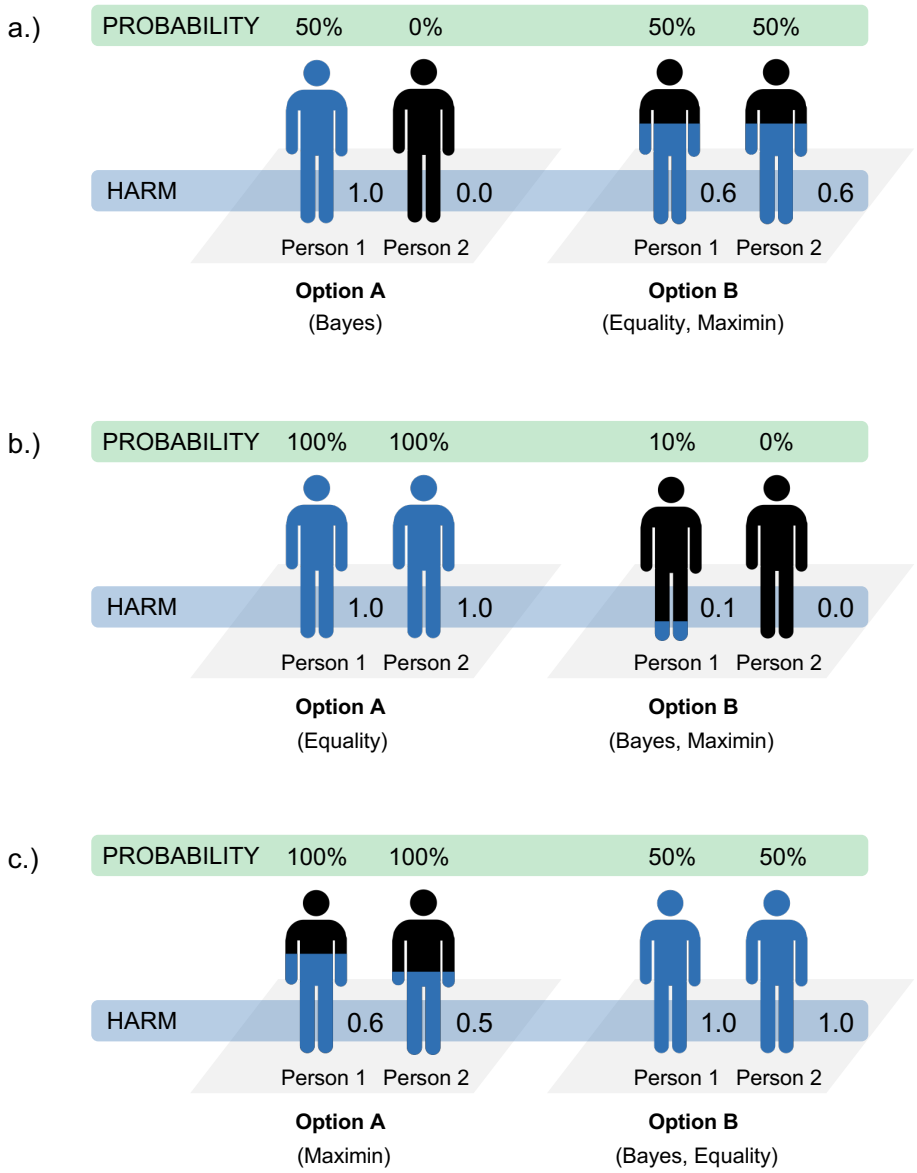


Figure 3.9: Utilizing distinct ethical principles for the allocation of risk yields varying outcomes. Three hypothetical scenarios, designed as simplified choices between two options (A and B), exemplify the trade-offs associated with these principles in risk distribution. In each scenario, two fictional individuals are assigned collision probabilities $p_{\text{collision}}$ and estimated harm H. While option A aligns with each of the three principles in every instance, option B may appear as an intuitive alternative to many, presenting compelling reasons to incorporate all three principles rather than relying on a single one. The figure was adapted from a previous author's publication [268].

Thus, these factors can be attributed to a road user's moral responsibility for the risks that arise. Consequently, the degree of moral responsibility that people bear for creating risky situations can be considered [66].

This approach promotes fairness and encourages responsible behavior among human road users. It addresses concerns regarding the potential misuse of AVs' crash avoidance systems, such as pedestrians purposefully exploiting the vehicle's ability to stop abruptly [5, 21]. Such irresponsible conduct can be discouraged by incorporating responsibility in risk distribution. This perspective aligns with the stance of the EU Commission, which advocates for fostering a culture of responsibility (R8) [11]. However, associating costs with irresponsible behavior could be perceived as a sort of punishment and be heavily debated. Another more positive view to look at it would be to frame the responsibility to avoid a collision, which is a basic principle of traffic rules. For example, the case of a right-of-way regulation requires road users to take responsibility for a possible collision. Thus, the implementation of these rules in AVs is necessary to participate in structured road traffic. Open remains the question of how to deal with rule breaks, i.e., misconduct.

Traffic rules are designed to facilitate safe interactions between road users, minimizing accidents and harm while ensuring traffic flow. They operate under the assumption that individuals can rely on others' adherence to these rules as long as they behave appropriately themselves. Therefore, it seems reasonable to account for any violations of traffic rules within the risk distribution framework. However, unlike the formal safety verification approaches discussed earlier, strict boundaries are not enforced in the proposed algorithm. Instead, rule violations are converted into costs associated with responsibility risks.

While various sets of rules can be incorporated into the motion planning framework, the proposed implementation calculates the responsibility of road users using the reachable set of each user [293]. A reachable set defines a time-variant area a road user can physically and legally reach according to traffic rules [292]. If a potential collision occurs within a road user's reachable set, they are considered not responsible for the collision. However, if a potential collision is outside of a reachable set, a collision is either not possible (because of the physical constraints), or the road user can be considered responsible, as shown in Figure 3.10. In such cases, they will be assigned costs associated with responsibility.

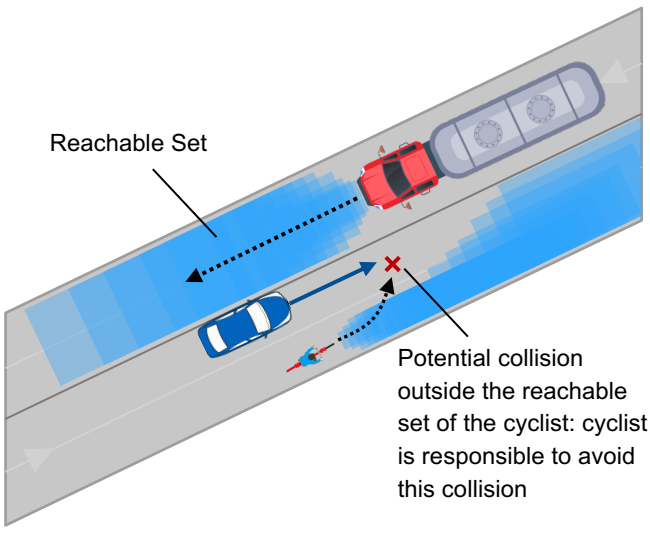


Figure 3.10: Schematic example for the responsibility principle in the case of a rule violation. A reachable set of all legal behaviors for all road users builds the basis for assigning responsibility. The figure was adapted from a previous author's publication [268].

In line with the principle of strict liability, the responsibility-based approach must also account for the characteristics of different road users and their corresponding moral claims. VRUs, such as cyclists or pedestrians, introduce significantly less risk to road traffic compared to cars or trucks due to their lower mass and velocities. If it weren't for motorized road users, the number of traffic fatalities would be significantly

reduced. Consequently, motorized participants have a special responsibility to protect weaker road users. Algorithmically, this is implemented by assigning a small responsibility value to motorized road users, irrespective of their behavior, but based on their mass and the associated risk they introduce into traffic.

It is important to allow for consideration of various degrees of responsibility: for example, not using the indicator seems less critical than running a red light. Therefore, responsibility costs J_R are calculated as a product of the individual risks of a road user R_i and the amount of responsibility according to the case r_i (Equation 3.16). The precise values for r_i representing the type and severity of the rule violation warrant further investigation and deliberation in future research, alongside determining weighting parameters for the Bayes, equality, and maximin principles.

$$J_R(u) = \sum_{i=1}^{|S_R|} r_i(u)\ R_i(u)\ ,\ r_i \in [0,\ 1) \tag{3.16}$$

In contrast to other risk costs discussed before, responsibility costs refer to a specific road user. Therefore, adjusting the risks for the responsible road user in the case of responsibility seems reasonable. Accordingly, the responsibility risk has the opposite sign to the other principles in the equation: in the case of responsibility, risk costs are subtracted, and thus, the risk of the responsible road user is weighted less. This model design bears substantial danger: In theory, the negative responsibility costs could exceed the other positive risk costs in the function. This would result in the AV deliberately trying to increase the risk for the responsible road user. This corresponds to punishing the responsible road user, which is not the aim of the proposed approach. The objective is to consider the responsibility in the event of a tradeoff in risk distribution. This is ensured once it is guaranteed that the risk of a road user never becomes negative in the risk cost function. This is achieved by replacing w_R by $-w_B$ normalized with $|S_R|$ as in Equation 3.17, and limiting r_i to less than 1 (Equation 3.16).

$$w_R = -\frac{w_B}{|S_R|} \tag{3.17}$$

Subsequently, Equation 3.18 represents the final risk cost function with the presented ethical principles.

$$J_{\text{Risk}}(u) = w_B\ J_B(u) + w_E\ J_E(u) + w_M\ J_M - \frac{w_B}{|S_R|}\ J_R(u) \tag{3.18}$$

4 The Ethical Vehicle Experiment

Chapter 3 proposed an algorithm for AV trajectory planning that considers various ethical principles for risk distribution. These ethical principles are integrated as top-down principles by means of a cost function that evaluates a sampled trajectory. However, in line with the requirements of Chapter 3, the weightings of these ethical principles are not yet fixed to eventually leave room for cultural or individual adaptation following a bottom-up approach. This chapter shows how these parameters can be determined according to Rawls' idea of a veil of ignorance. Therefore, this chapter will present a user survey to capture the moral views of society. The core of this chapter is to establish a method that enables a connection between a user survey and the proposed algorithm so that the survey results and the according user opinions can be integrated within the algorithm. This study is conducted to determine values for w_B, w_E, and w_M. The underlying method can be transferred to find additional weight parameters, e.g., in the context of the responsibility principle. The results of the study are shown in Section 5.1, while further discussions on shortcomings and validity within the overall context of trajectory planning and decision-making are provided by Chapter 6.

The user study is publicly available as an online experiment named *TUM Ethical Vehicle Experiment* and can be accessed via: *https:// ethicalvehicle.ftm.mw.tum.de/*

4.1 Motivation & Goal

The ultimate goal of this study is to calculate three values for the parameters w_B, w_E, and w_M for each individual user based on their answers in the survey. On the one hand, this requires sophisticated question-answer pairs to provide detailed information and calculate the resulting parameters. On the other hand, the questions should be easily accessible to every survey participant. The study should not include complexity barriers that could potentially prevent the whole of society from participating. This trade-off must be explicitly taken into account for the study design. To create a reasonable database, as many participants as possible should be approached but without the need for financial incentives for participation. Although criticized for focusing on discriminatory factors [118], the Moral Machine Experiment [70] gained much attention with more than 40 Mio. participants. This goes in parts back to the survey questions, some of which raised dubious moral conflicts. However, the experiment also became widespread because the users were presented with their results at the end and could compare themselves with other study users. This kind of reward for survey participants should serve as an example here. Table 2.1 presented the number of participants in existing studies, which range from 32 to 40 Million with a median of 182. Therefore, evaluating the results on the basis of around 200 participants is set as a goal for this study.

4.2 Study Design

The Ethical Vehicle Experiment consists of four steps, shown in Figure 4.1. At the beginning of the study, participant metadata is requested, such as gender, age, and nationality. Since the target group is not controlled in a publicly accessible study, the metadata should help to describe the group of participants and put them in relation to a representative target group. In the next step, the experiment is explained in detail, presenting an exemplary question to the user. At the end of the explanation, the user is asked if the task and the corresponding questions were understood. Participants gain access to the survey only if they declare that everything is clear to them. Otherwise, they are asked to provide more information on what is unclear. If the survey participant understands the task, nine binary questions are asked. This method was used to increase the quality of the study instructions through pre-tests until no more uncertainties remained on the part of the participants. The questions, as well as the answer options, are randomly permuted in order so that any bias in terms of ordering is mitigated. Section 4.2.1 elaborates on how these questions are generated to establish a connection to the proposed model from Chapter 3. Finally, after answering these questions, weightings are calculated as a result and presented to the user compared to the previous participants, similar to the Moral Machine Experiment [70]. Section 4.2.2 provides more details on how to calculate the weighting parameters as a result of the study. In addition to the metadata and the results, each user's time to answer the questions is recorded to avoid samples in the database, where users clicked randomly through the study. Participation in this study usually takes around 5-10 minutes.

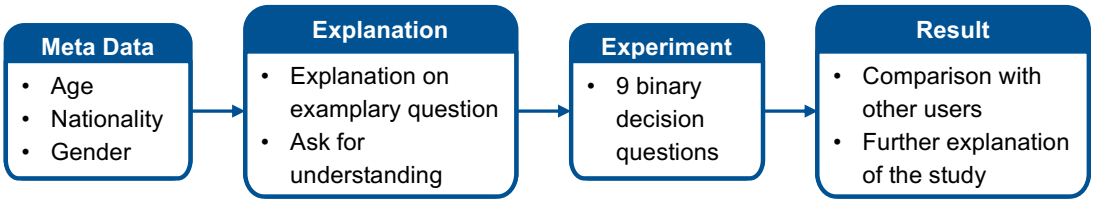


Figure 4.1: Procedure of the user study in four steps: collection of the metadata, explanation of the study, answering the questions, and presenting the results to the participant.

As stated, the participants are asked nine questions during the experiment. Each of the questions represents a fictive scenario. The following describes these fictive scenarios in a similar way as they are presented to the user in the explanation part of the experiment.

Since the problem of fair risk distribution can initially be considered independently of the application of autonomous driving, the scenarios are abstracted from that particular use case to a more general view of the problem of risk distribution. In addition, providing only the essential information keeps the experiment simple and accessible. As a result of this abstraction, each scenario presents two persons, as shown by Figure 4.2: There are two options with different risk distributions for each person. The user must choose one option, A or B, that better reflects their moral views regarding risk distribution. The users themself do not participate in the fictive scenario but observe the situation from an outside perspective behind a *veil of ignorance*.

The risk conditions for each person in an option are described by two values: First, the collision probability from 0 % to 100 % indicates the likelihood that a person will receive harm, as introduced in Chapter 3. Second, following the risk definition from Equation 3.1, the estimated harm out of that collision is the second dimension taken into account here. According to the algorithmic implementation, this value ranges from 0 to 1, where 0 means no injury and 1 stands for the greatest possible injury. Analogous to the algorithm, harm only refers to personal injury and does not take into account any property damage. As shown in Figure 4.3 the collision probability is assigned as a value right above the person symbol as a percentage value. Harm is visualized by filling the visualized person with a blue color with respect to the vertical axis.

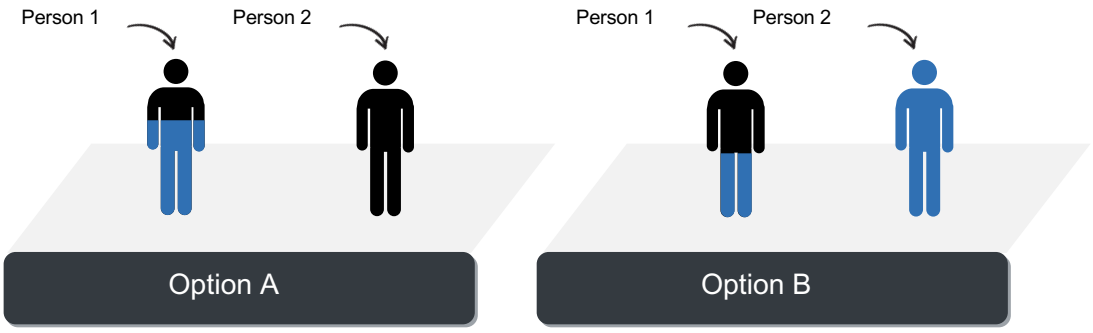


Figure 4.2: Each question of the survey is represented by a fictive scenario consisting of two options with two persons that have different risk conditions.

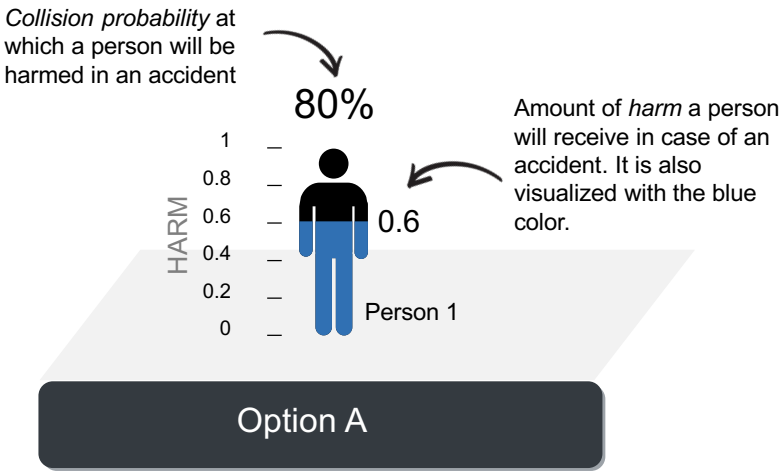


Figure 4.3: A risk condition for a person is described by its collision probability and the according harm as visualized here.

Putting these things together, Figure 4.4 presents an example question as asked in the experiment. Two persons have different risk conditions in two options, A and B. The questions are constructed so that each option represents at least one of the ethical principles. In this case, person 1 will receive a harm of 0.6 with 80 % probability in option A. This implies that with a 20 % probability, person 1 will be unharmed. In option B, person 1 will receive a harm of 0.4 with 70 % probability. Person 2 will definitely be unharmed in option A while receiving the greatest possible harm (1.0) with 20 % probability in option B. The survey participant must make a decision in each case, and the questions cannot be skipped. Visualization of the probabilities ("Show Visualization") and a text-form description for the two options ("Show Description") provide additional support.

The visualization should facilitate an evaluation of the distribution of risk upon first observation without considering any numerical values. It involves depicting ten fictitious persons for each person and option. Depending on the probability, these fictitious persons are harmed (indicated by blue coloring) or remain unharmed. For example, to reflect a probability of 20 %, two of ten persons are assigned with the associated harm, while the remaining eight are displayed as unharmed. To understand the probability and outcome of the accident for a particular person, one can randomly select one person from the ten fictitious persons. This person's outcome corresponds to the probability and outcome of the accident for that particular person.

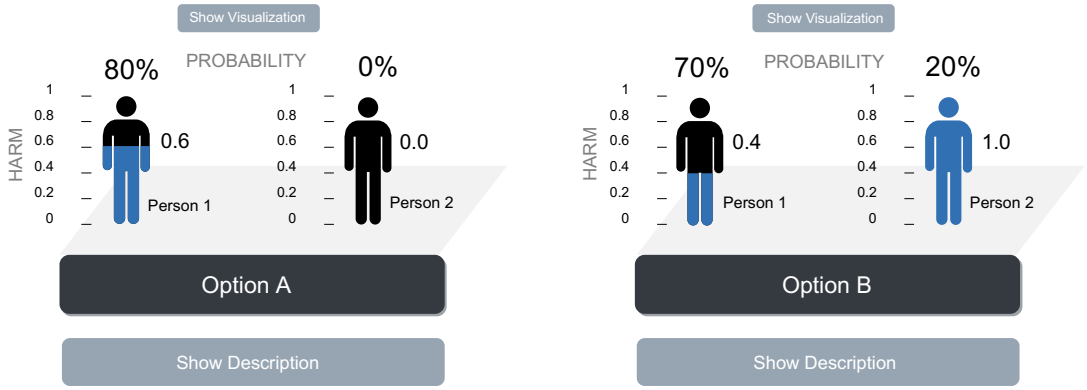


Figure 4.4: Examplary question as asked in the experiment with two options for two persons and their visualized risk conditions.

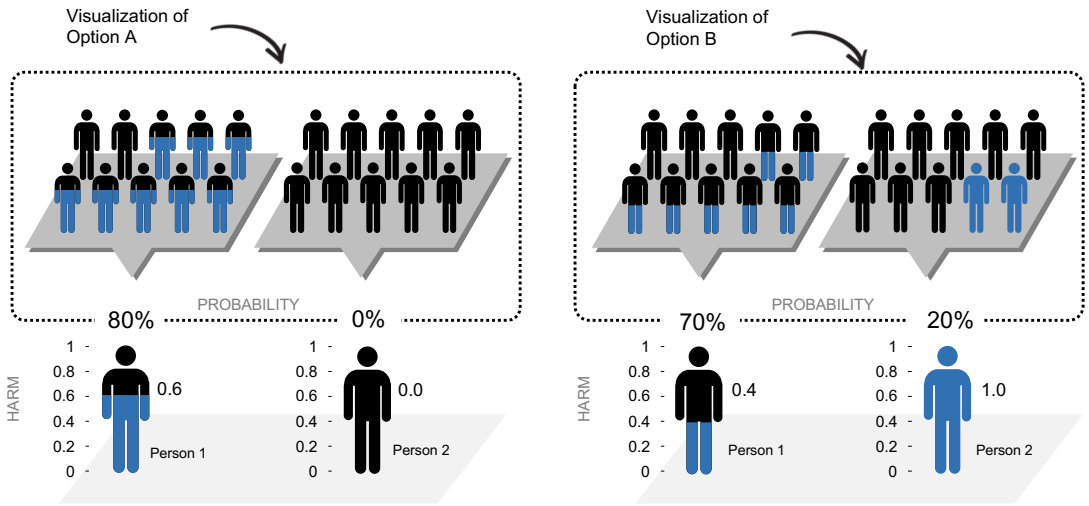


Figure 4.5: The probabilities are further visualized by depicting one person as a group of ten, analogous to an urn problem.

After these instructions, the participant is presented with nine questions all following the scheme of Figure 4.5 with varying risk conditions for the depicted persons. All nine questions with the risk conditions as derived by the following Subsection 4.2.1 can be found in Appendix A. After answering the questions, each user is presented with their result compared to the previous users, as shown by Figure 4.6. Therefore, the weightings for the utilitarian (=Bayes), equality, and maximin principle are visualized. In contrast to this work, the Bayes principle is referred to as the Utilitarian principle since this term is more common outside of academia. In addition, the user gets some further explanation of these principles and their underlying theories. Finally, the experiment also reflects the user on how well their decisions can be mapped using a linear combination of these three principles. It may happen that the underlying model consisting of the three ethical principles cannot represent the user decisions well, regardless of the parameterization. In this case, the participants are asked to provide some further insights into how they made their decisions.

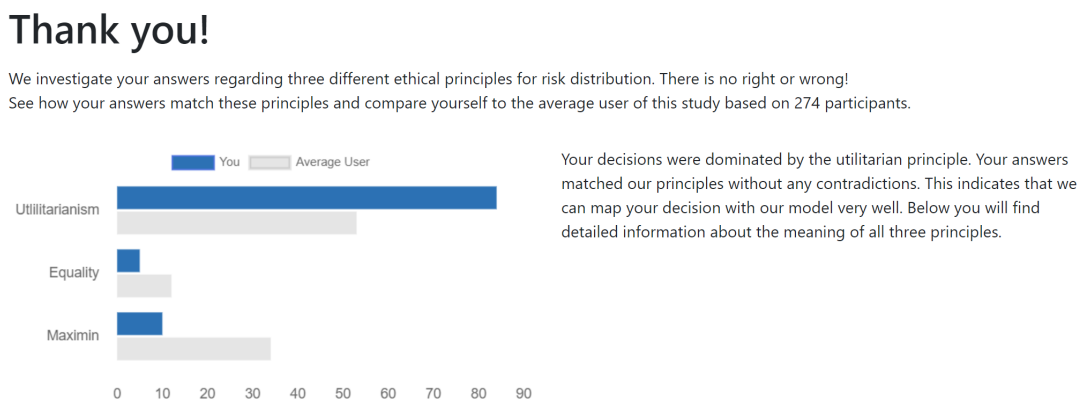


Figure 4.6: The result presentation to survey participants includes a comparison of the user's parameters (blue) with the average user (grey), as well as some further information regarding the ethical principles and compliance with the underlying model.

4.2.1 Survey Question Generation

As motivated in the previous section, the study's goal is to determine the personal inclination of survey participants with respect to the distribution principles according to Bayes, equality, and maximin. By answering a number of questions, it should be possible to determine a set of weighting parameters $\boldsymbol{w} = [w_B, w_E, w_M]$ for each user that describes the weightings of the respective principles according to their moral views. The questions should be as simple and accessible as possible but, at the same time, reflect the underlying aspects of risk distribution.

The simplest question design results in two options for the user. Each option must have at least two persons with different risk loads to face the aspect of risk distribution. Finally, to account for the two-dimensionality of risk, each person in each option must be assigned a value for probability and a value for the expected harm. This results in $2^3 = 8$ parameters that define a single question as the minimum to account for the underlying model. Since the distribution principles are formulated independently of the number of road users involved (Equations 3.13-3.15), the risk distribution can be extrapolated from this simple case to more complex cases with more road users. For each option A and B in a scenario, the cost terms J_B, J_E, and J_M can be calculated respectively using Equations 3.13-3.15. Table 4.1 gives an overview of the parameters and costs used in a question.

Table 4.1: Notation of the relevant parameters for the study, which consists of two answer options (A and B) and two persons (1 and 2) with varying collision probability and estimated harm. This results in different costs J_B, J_E, and J_M according to the ethical principles.

	Option A		Option B	
	Person 1	Person 2	Person 1	Person 2
Collision probability p	$p_{1,A}$	$p_{2,A}$	$p_{1,B}$	$p_{2,B}$
Estimated harm H	$H_{1,A}$	$H_{2,A}$	$H_{1,B}$	$H_{2,B}$
Weighted Bayes cost J_B	$w_B J_B^A$		$w_B J_B^B$	
Weighted equality Cost J_E	$w_E J_E^A$		$w_E J_E^B$	
Weighted maximin cost J_M	$w_M J_M^A$		$w_M J_M^B$	

For both options A and B, the risk-cost function (Equation 3.11) can be calculated with the weights w_B, w_E, and w_M as variables. The survey participant's choice between A and B indicates which of the two costs is perceived to be lower. If, for example, the participant favored option A, then the costs for option A must be

lower than those for option B according to their moral views. Consequently, a question q with its answer can be formulated by the inequality I_q given by Equation 4.1, where the user's choice of an option determines the direction of the inequality sign. Due to the exclusively linear correlation, Equation 4.1 can be simplified to Equation 4.2, where the weights depend only on the differences of the costs. According to the underlying model, a question is characterized by the three values $\boldsymbol{q} = [\Delta J_B, \Delta J_E, \Delta J_M]$. These three values define a question that can be formulated as inequality depending on the user's choice. Note that, due to the linear model, several possible questions with different risk conditions hold the same information from the model's perspective.

$$I_q : \begin{cases} w_B J_B^A + w_E J_E^A + w_M J_M^A < w_B J_B^B + w_E J_E^B + w_M J_M^B & \text{if user choice is option A} \\ \\ w_B J_B^A + w_E J_E^A + w_M J_M^A > w_B J_B^B + w_E J_E^B + w_M J_M^B & \text{if user choice is option B} \end{cases} \tag{4.1}$$

$$w_B \Delta J_B + w_E \Delta J_E + w_M \Delta J_M < 0 \quad \text{with} \quad \Delta J = \begin{cases} J^B - J^A & \text{if user choice is option A} \\ J^A - J^B & \text{if user choice is option B} \end{cases} \tag{4.2}$$

The approach to question generation is to restrict the solution space of all possible weighting triplets by answering binary questions. Therefore, a system of inequalities $I_{q1}, ...I_{q9}$ should be set up that can be translated into questions. To define these inequalities, the solution space of all possible sets of weighting parameters $\boldsymbol{w}$ is examined with the constraint $w_B + w_E + w_M = 1$ from Chapter 3. The set W describing the solution space is thus obtained as follows:

$$W = \{w \in \mathbb{R}_0^{3+} | w_1 + w_2 + w_3 = 1\} \tag{4.3}$$

The solution space W can be represented by a three-dimensional triangle between the axes w_B, w_E, and w_M, as shown by Figure 4.7. Each inequality (Equation 4.2) representing a question can be interpreted as a space that intersects the solution space W accordingly. In Figure 4.7, an exemplary question is simplified as a cut through the solution space. Each side of the straight intersection refers to one answer option in the question and reflects the solution space that aligns with the answer to that question.

For formulating a set of questions, the objective is to identify n intersecting straight lines that divide the solution space W, where the value of n corresponds to the number of questions. The number of questions is a design parameter of the study. Given the emphasis on maximizing user outreach without introducing additional incentives, n is set to nine. Due to the three underlying principles and for the reason of symmetry, it is advantageous to choose n so that it can be divided by three. The further requirement of an easily accessible study suggests that the parameters should only differ in the first decimal digit to avoid complex numbers. This requires that the values for ΔJ are also integer multiples of 0.1. Another requirement is that the solution space is divided symmetrically to avoid bias against any principle.

To generate nine questions, only three inequalities are needed with the aim of dividing the solution space as homogeneously as possible. Out of each inequality, three questions can be derived by permutation according to the ethical principles. The requirement of integer multiples of 0.1 results in a manageable number of possibilities. Note that each multiple of the equation represents the same intersection line and thus results in the same inequality I. Equations with large values in ΔJ are preferred for generating significant questions

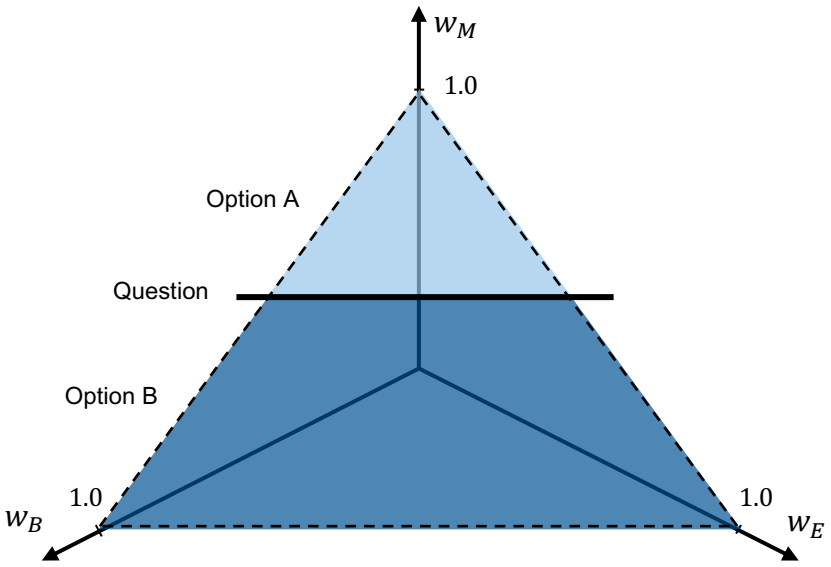


Figure 4.7: A question can be interpreted as constraining the solution space W, whereas each side of the straight cut can be related to one option as an answer to the question.

regarding the selected ethical principles. The larger the values are for ΔJ, the larger the differences in costs regarding an ethical principle, and the better a trade-off becomes apparent.

The inequalities and, thus, questions arising from these requirements and principles are shown in Figure 4.8. The corresponding value triplets for ΔJ representing a question are depicted by Table 4.2. Next to the requirement of simple numbers and large differences, these values have been chosen so that they result in a discretization grid as homogeneous as possible. Figure 4.8 shows that this is only possible to a limited extent. The solution space between $q2$, $q8$, and $q9$, for example, is significantly larger than that between $q2$, $q7$ and $q8$.

Table 4.2: Design parameters for the nine questions in the Ethical Vehicle Experiment. A positive value indicates higher costs for a specific principle in option A than in option B. For example, $q1$ has the same Bayes cost J_B for each option ($\Delta J_B = 0.0$), but equality costs J_E are higher by 0.4 ($\Delta J_E = 0.4$) in option A and maximin costs are lower by 0.4 ($\Delta J_M = -0.4$) than in option B.

Question No.	ΔJ_B	ΔJ_E	ΔJ_M
$q1$	0.0	0.4	-0.4
$q2$	0.1	0.1	-0.5
$q3$	0.1	-0.5	0.1
$q4$	0.3	0.3	-0.3
$q5$	0.3	-0.3	0.3
$q6$	0.4	0.0	-0.4
$q7$	0.4	-0.4	0.0
$q8$	-0.3	0.3	0.3
$q9$	-0.5	0.1	0.1

4.2.2 Result Calculation

Given a set of inequalities denoted as $\{I_{q1}, ..., I_{q9}\}$ arising from the responses of survey participants, the participant's specific weight configuration $\boldsymbol{w} = [w_B, w_E, w_M]$ should be calculated. A system of n inequalities emerges from the questions. These inequalities do not necessarily have to be consistent. In particular, this

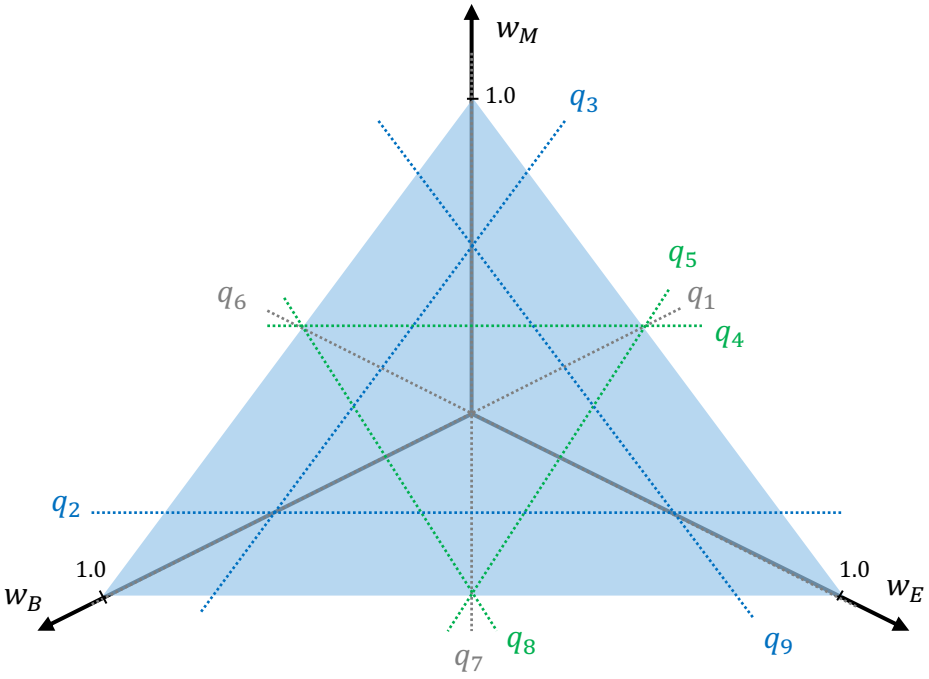


Figure 4.8: The user survey consists of nine questions $\{q1, ..., q9\}$ that are intended to divide the solution space (blue triangle) as equally as possible while maintaining comprehensible parameters in the user study. The colors indicate question triplets that result from permutation.

can happen for two reasons: either the underlying principles do not reflect the moral views of the participant, or assuming that the principles are a proper reflection, the participant's responses are internally inconsistent. This situation may lead to the absence of a solution that simultaneously satisfies all the inequalities.

Therefore, to find an appropriate solution $\boldsymbol{w}$, a loss function quantifies the violation of an inequality with respect to a configuration $\boldsymbol{w}$. The violation is assumed to be stronger when the distance between the configuration point and the inequality straight is higher. According to Equation 4.4, the loss for a given configuration $\boldsymbol{w}$ is zero for $\boldsymbol{w}$ if it fulfills the inequation. In the case of a violation of an inequality, the loss equals the Euclidean distance between $\boldsymbol{w}$ and the inequality I. Since there can be multiple violations of inequalities at the same time, the loss of a given solution candidate $\boldsymbol{w}$ is finally the sum of distances over all the questions.

$$\mathcal{L}(\boldsymbol{w}) = \sum_{i=1}^{|Q|} \begin{cases} 0 & \text{if} \quad \boldsymbol{q_i w}^T < 0 \\ \frac{\boldsymbol{q_i w}^T}{\|\boldsymbol{q_i}\|} & \text{otherwise} \end{cases} \tag{4.4}$$

A total loss of zero implies that, for the set of nine questions, the model formulated based on the Bayes, equality, and maximin principles can accurately represent the participant's preferences. On the contrary, a higher total loss indicates a poorer fit between the model and the participant's ethical choices. This metric will serve as a means of validating the proposed model in Chapter 6.

To calculate the parameters $\boldsymbol{w}$ according to the user's choices, the loss function defined in Equation 4.4 is minimized as shown in Equation 4.5. There are several possible methods to solve this optimization problem described by the optimization objective 4.5 and the inequalities as constraints. Due to the fact that consistency cannot be guaranteed, it has been shown to be beneficial to discretize the solution space (e.g., with 0.01) and to calculate the loss values accordingly across the whole solution space to generate a loss map. This method provides fast computation times and is accurate enough considering the rough discretization of the solution spaces. In case multiple points yield the same loss of zero, the centroid of these equivalent points is used as the result.

$$\min_{w \in \boldsymbol{W}} (\mathcal{L}(w)) \tag{4.5}$$

5 Results

Chapter 3 proposed an algorithm with integrated ethical principles for AV trajectory planning, while Chapter 4 presented a method to determine the algorithmic parameters. This chapter will show the results for both methods as the main component of this thesis. Since the results of the user study from Chapter 4 provide the relevant parameters for the algorithm from Chapter 3, Section 5.1 will first present the results of the user study denoted as Ethical Vehicle Experiment. In particular, the resulting weighting distributions (Section 5.1.1) and the conformity of the participants with the ethical principles (Section 5.1.2) will be illuminated. Section 5.2 will then analyze the parameterized algorithm in an empiric simulation. First, the effects of each of the ethical principles will be investigated in Section 5.2.1 to provide some further insights into the effect that those principles have in trajectory planning. Second, the parameterized algorithm as the proposed approach to considering ethics in trajectory planning will be compared to a state-of-the-art algorithm and a version with selfish behavior, as motivated in the introduction. It will provide some further insights into the algorithm by analyzing correlations and, thus, actual trafe-offs between the ethical principles. Finally, in Section 5.3, the proposed idea of maximum acceptable risk from Section 3.3 is added to the ethical risk distribution and empirically evaluated based on the simulation.

5.1 Ethical Vehicle Experiment

The Ethical Vehicle Experiment as presented by Chapter 4 was conducted to determine values for the weighting parameters w_B, w_E, and w_M. The URL to the online survey was publicly available and shared via social media (Linkedin) and in an article at Nature Machine Intelligence [268]. A total of 273 participants took part in the experiment. Before evaluation, samples that are not meaningful should be filtered. In particular, there have been participants who apparently randomly clicked through the survey to access the results at the end. Therefore, the time required to understand and assess each question and make a moral choice is assumed to be at least two seconds. So, in the first preprocessing step, samples were neglected, in which at least one question was answered in less than two seconds. This leads to 247 participants, on which the results are reported in the following. Since the experiment was mainly advertised via social media, the group of participants is biased. Figure 5.1 shows the composition of the participants in terms of gender, age, and nationality. The majority of the participants were men aged 20 to 30 years. Evaluation of the nationalities shows a clear bias towards Germany, which indicates an overrepresentation of a particular subgroup. Females and people over 40 years of age are underrepresented in the study. The results presented below must be understood in this context.

5.1.1 Weighting Parameters

The parameter configurations w were calculated for each participant in the experiment according to Section 4.2.2. This results in 247 parameter triplets for the Bayes, equality, and maximin principles, as shown in Figure 5.2 according to the number of participants. The size of the blue balls in the diagram corresponds to the number of users who prefer this specific configuration. As can be seen, most preferences accumulate at

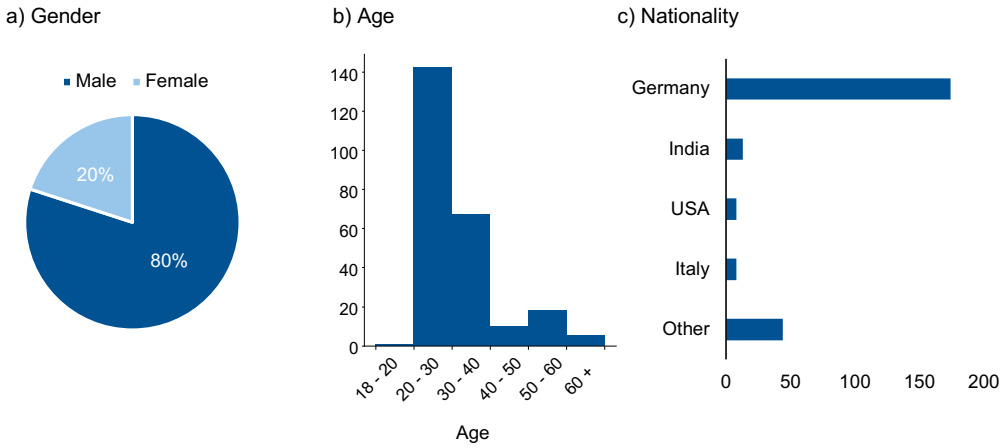


Figure 5.1: Distribution of the survey participants regarding a) gender, b) age, and c) nationality

low values smaller than 0.2 for the equality principle. Hence, the Bayes and maximin principles dominated the users' decisions. It is also apparent that no specific configuration of w is particularly dominant, but there is a distribution of different configurations in this area.

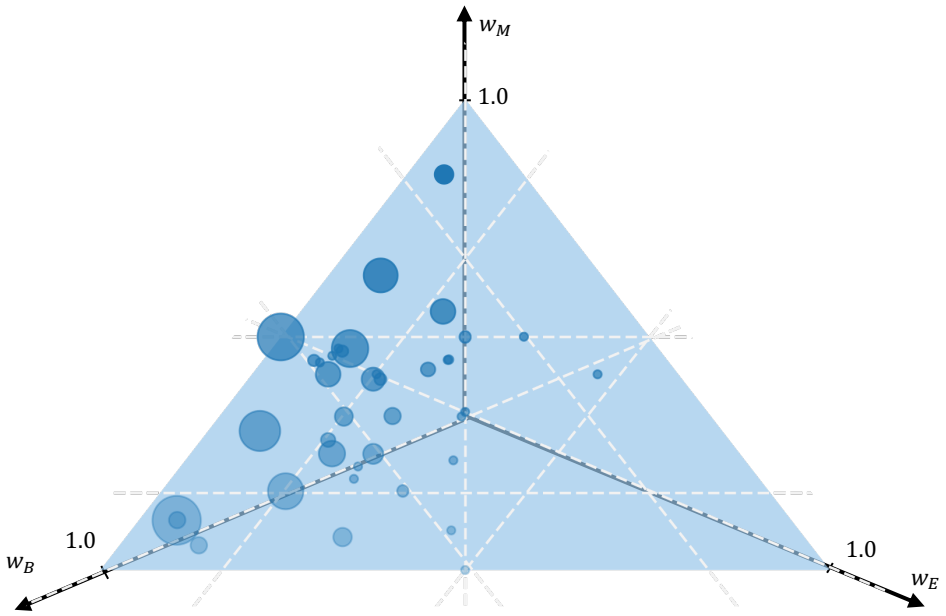


Figure 5.2: Results from the Ethical Vehicle Experiment shown in the solution space. The larger the blue balls, the more participants are reflected by a specific configuration.

The distribution of the weighting factors according to the specific principles reveals further insights. Figure 5.3 shows boxplots for the weightings according to Bayes, equality, and maximin. Looking at the mean values, the Bayes principle dominates with 53.5 %, followed by the maximin principle with 34.7 % and the equality principle with 11.7 %. However, it is worth noting the high deviations, which, in the case of the maximin principle, for example, extend over about 85 %-points. Only in the case of the equality principle is the variance lower, at around 35 %. In order to parameterize the algorithm, the corresponding mean values of the principles are used in the further course of Section 5.2. The extent to which this corresponds to the objective of a social consensus against the background of this data will be part of the discussion in Chapter 6.

The evaluation of how the age or gender of the participants impacts the result led to no further insights. Either the distribution is comparable across all relevant subgroups of participants (e.g., between males and females), or the number of representatives for a subgroup is too small to draw reliable conclusions.

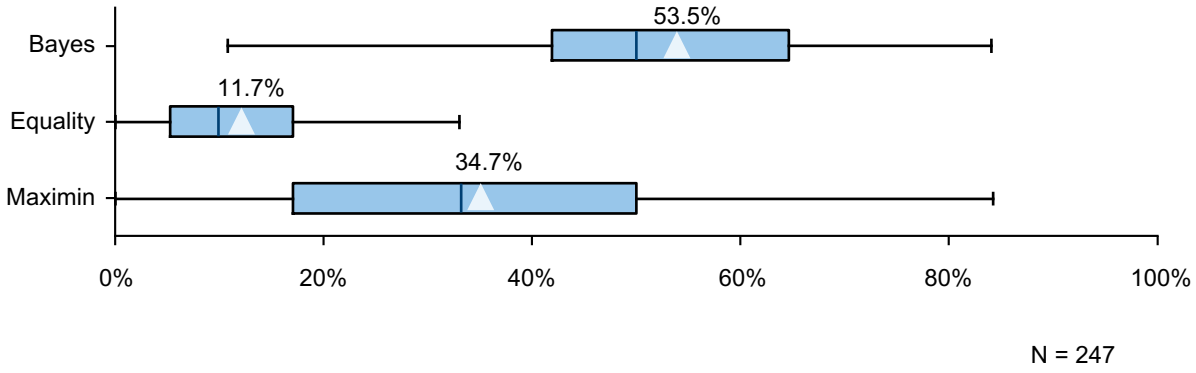


Figure 5.3: Distribution of the ethical principles displayed as boxplots as the results from the Ethical Vehicle Experiment. The data are presented as median (Q2) between Q1 and Q3 with horizontal lines. The mean is marked with a triangle, and the whiskers are Tukey style.

5.1.2 Model Validation

In addition to evaluating the configurations of w of the various participants, it is also worthwhile to look at the compliance of the users' moral views with the underlying model. Therefore, the loss $\mathcal{L}$ according to Equation 4.4 is evaluated for each user. Figure 5.4 shows this loss distribution among the survey participants as a measure for contradictions with the underlying model. The higher the loss, the more the moral views of the participants conflict with the linear combination of the three proposed ethical principles. Among all participants, the model can represent the answers to the nine questions of the experiment for $80.6\,\%$ of all participants ($\mathcal{L} = 0$). To put this into context, if the questions are answered randomly, the probability that the model can resolve these answers without contradiction is about $7\,\%$. The mean value of $\mathcal{L}$ when answering the questions randomly is 0.47. If the value for $\mathcal{L}$ is greater than 0.2, this is seen as an indicator of poor representation of the users' moral views by the model. This applies to $11.7\,\%$ of the participants. In that case, at the end of the experiment, the participants were asked to provide further explanations on how they made the decisions. However, there was no feedback from the 29 participants concerned in this regard, so no further conclusions regarding their guiding principles can be drawn at this point. A share of about $2\,\%$ (five participants) shows strong deviations from the model with $\mathcal{L} > 0.6$.

5.2 Risk Distribution Principles

The results of the Ethical Vehicle Experiment shall next be used to empirically examine the algorithm proposed in Chapter 3. The following Section 5.2.1 places particular emphasis on the three distribution principles denoted as Bayes, equality, and maximin principle. Section 5.2.2 investigates the ethical approach to trajectory planning as a combination of ethical principles using the parameters of Chapter 4 in contrast to a selfish algorithm variation as motivated in the introduction. The interactions between these principles against the background of trajectory planning are further investigated in Section 5.2.3 by means of a comprehensive correlation analysis.

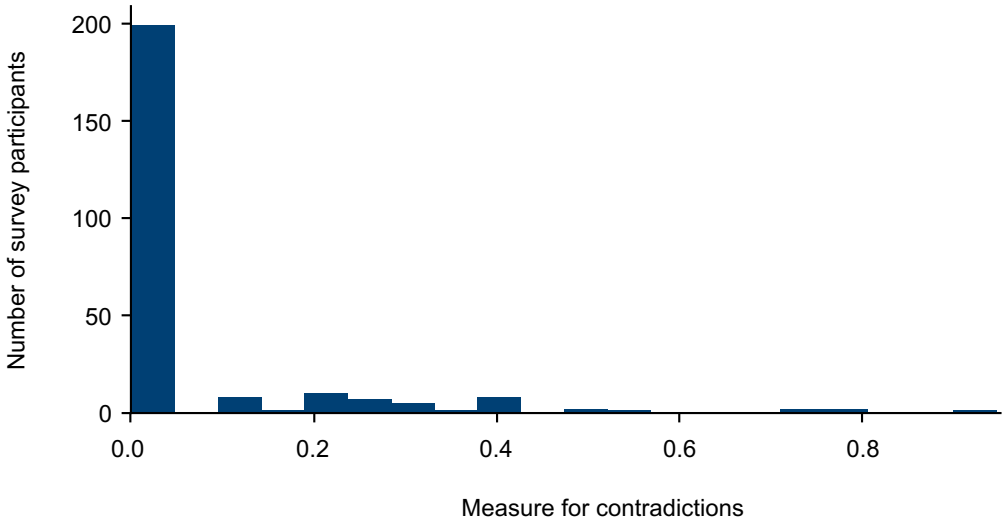


Figure 5.4: Histogram of loss values for the participants of the Ethical Vehicle Experiment. The loss value can be seen as a measure of contradictions. The majority of participants achieved loss values close to zeros, indicating a proper model representation.

The empirical evaluation of the algorithm is based on a simulation of 2,000 scenarios with the platform CommonRoad [294]. Each CommonRoad scenario provides a planning problem consisting of an initial state for the AV and a goal state. A combination of diverse variables, such as a local area, a time horizon, or a target velocity, defines the goal state. The road networks are described by lanelets [277] and are either extracted from real-world road networks or crafted by hand. This results in diverse scenarios from different countries, such as Germany, the USA, or China, and a diverse set of situations, such as cities, country roads, and highways.

Figure 5.5 shows an exemplary hand-crafted scenario with a planning problem. Traffic in the scenarios is deterministic, either recorded or created by hand [295]. The hand-crafted scenarios are deliberately created as critical scenarios [296, 297], which can be understood as edge cases and are hard to solve in terms of trajectory planning. The results of the CommonRoad Motion Planning Competition 2023 with solution ratios from $36\,\%$ (Phase 1) [298] to $50\,\%$ (Phase 2) [299] from the winning algorithms underline the difficulty of these scenarios. As a consequence, the CommonRoad scenarios do not serve as a representation of real road traffic and higher risks than in road traffic can be expected here. The total number of all scenarios requires the planning of around one million trajectories if there are no accidents that terminate a scenario. For the evaluation of the various trajectory planning algorithms, the presented prediction algorithm *Wale-Net* is used to obtain probabilistic trajectory predictions, as shown in Figure 5.5.

The deterministic behavior of road users in the CommonRoad scenarios brings some special characteristics that need to be taken into account. On the one hand, this determinism results in a high degree of comparability, as no dependencies and interactions with behavior models influence the results. The fact that the AV must react to all road users and cannot assume any reaction initially increases the difficulty of the planning task. This can lead to situations that are not solvable without a collision. While this aspect may not align with reality, it does not impede the assessment of ethical principles. On the other hand, this does not reflect reality well, as road users would usually react to AV behavior. So, situations can occur that would likely not occur in real life, such as a vehicle slowly rear-ending the AV. This is especially problematic for the responsibility principle introduced in Section 3.4.4, since here, the interaction with other road users is explicitly described by assigning responsibility. In the CommonRoad simulation, however, the responsibility for avoiding collisions always lies with the AV due to the determined behavior of the other road users. For this reason,

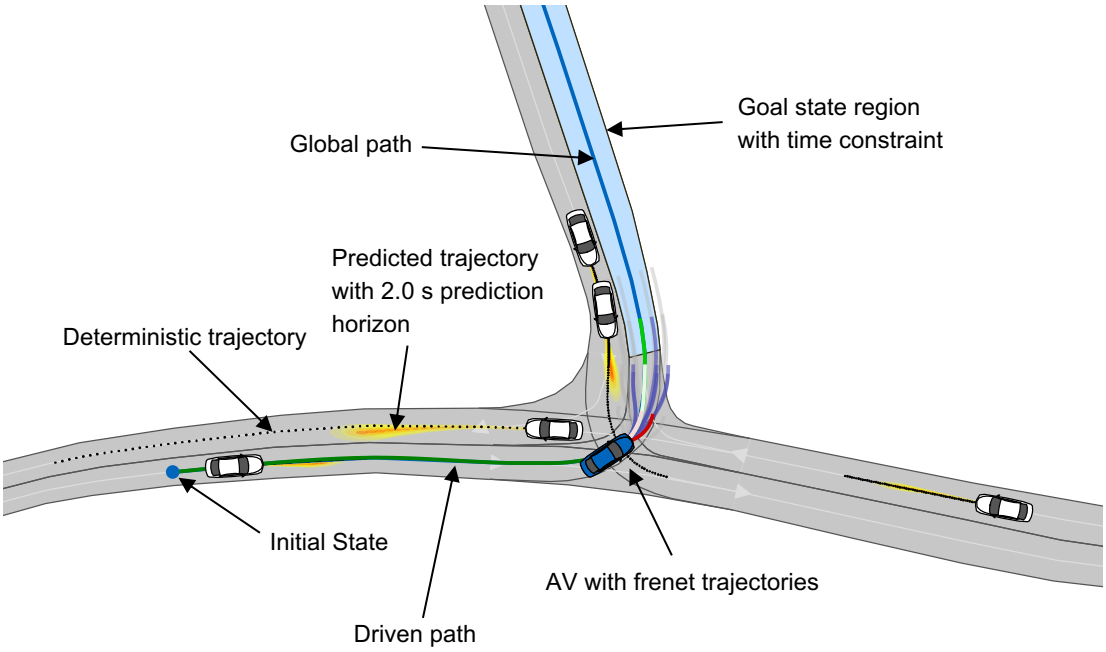


Figure 5.5: Exemplary hand-crafted CommonRoad scenario (ZAM_Tjunction-1_486_T-1) of an unprotected left turn with an initial state and a goal region as a planning problem. Due to the deterministic trajectories, the blue AV must turn left between two oncoming vehicles not to be rear-ended by the vehicle behind.

the responsibility principle cannot be meaningfully examined here, and the focus is placed on the remaining principles.

To be as independent of parameters as possible in the results, only the essential trajectory costs are taken into account in the cost function of the trajectory planner. In addition to the costs for risk, which represent the focus of the investigation, costs for the lateral distance J_{LC} to the global path are taken into account, as well as velocity costs J_{V} (Equation 5.1). The velocity costs amount to the difference between a current or planned velocity and a target velocity, which is either specified by the scenario or can be derived from the planning problem so that the goal is reached in the specified time. Cost terms relating to comfort are not taken into account, as evaluating the trajectories' comfort is outside the scope of this thesis. The weighting parameters w_{LC}, w_{V}, and w_{Risk} have not been optimized but were experimentally determined as part of a baseline algorithm and were fixed for all the experiments. The systematic optimization of such parameters is an additional field of study and has been shown to require lots of computational power and data [300].

$$J_{\text{total}} = w_{\text{LC}} J_{\text{LC}} + w_{\text{V}} J_{\text{V}} + w_{\text{Risk}} J_{\text{Risk}} \tag{5.1}$$

The vehicle model used in the simulation is a simple point mass model with acceleration limits and steering angle limitations. To focus on the planning of the trajectories, it is assumed that there are no control deviations between the planned trajectory and the trajectory actually driven.

5.2.1 Bayes, Equality and Maximin Principle

The Bayes, equality, and maximin principles will first be examined in separate algorithms to better understand the effects of the various ethical principles in trajectory planning. Therefore, three algorithms with the parameters given by Table 5.1 with the according names Bayes, equality, and maximin are evaluated in the simulation. The principles will thus be examined for their suitability for use in trajectory planning.

Table 5.1: Overview of algorithm parameters for Bayes, equality, and maximin algorithm

Algorithm	w_B	w_E	w_M
Bayes	1.0	0.0	0.0
Equality	0.0	1.0	0.0
Maximin	0.0	0.0	1.0

The respective cost functions with the corresponding parameterization represent a target prioritization but do not yet guarantee that the principles are also reflected in the actual risk distribution. This is due to two relevant effects in AV trajectory planning: Firstly, the calculated risks are probability-based assumptions about the future, which can accordingly turn out to have varying degrees of accuracy over time. The better the risk estimation and the more data or scenarios are evaluated, the less this effect should be. Secondly, the scenarios and the consideration of physical conditions represent restrictions that do not allow for distributing risk arbitrarily in road traffic. For example, decreasing the AV velocity to reduce the potential harm and risk to a VRU may also lead to a lower potential harm and risk for other road users around.

In the following, the risk distributions, as well as the actual harm resulting from simulated collisions, will be analyzed according to different road user groups. To examine the distribution aspect, a distinction is made between the ego-AV and all other road users (third party). In accordance with the EU regulation on demanding special attention to VRUs [139], this group will also be examined separately, denoted as VRU. Most of the risks that occur in the simulation are small and close to 0. In order to address the differences between the algorithms in the evaluation and to make them clear, not all risks are presented, but the 100 highest risks per algorithm are compared. The higher risk values are also more relevant for our evaluation purpose than the lower ones. This should serve as an indicator for risk distribution aspects. Increasing from 100 to $1,000$ or $10,000$ highest risks as evaluation ground shows the same effects but is less visible, which is why the 100 highest risks are shown in the following boxplots.

Figure 5.6 shows that, in general, for all three algorithms, risks are lowest for the ego-AV. The differences between the three algorithms and the spread of risks are significantly smaller than for the other road user groups. Nevertheless, the algorithm with the maximin principle results in the highest risks for the ego-AV compared to the other two principles. The same can be seen in a more prominent form among third-party road users and VRUs. Here, significantly higher risk values of over 0.3 are achieved with the maximin algorithm. Similar distributions between third-party and VRUs show that with all three algorithms, the highest risks of all third-party road users are dominated by VRUs. Since only the highest risks are considered here, this representation does not serve as the only indicator to assess the distribution of risk between the different user groups. For example, one would expect the Bayes algorithm to have the lowest overall risks, which is not the case according to Figure 5.6. This is because the Bayes principle, unlike the equality principle, allows for high individual risks as long as the total is low. Therefore, consideration of the highest occurring risks alone is not sufficient.

The occurring risks should be an indicator of the amount and severity of accidents that appear with a respective planning algorithm. However, this requires precise knowledge of risk with high accuracy. To validate the relationship between the calculated risks and the actual accident outcome, Figure 5.7 shows the cumulated harm in over 2,000 scenarios for all three algorithms. It is noticeable that the cumulative harm for the ego-AV is significantly higher than the previously calculated risks compared to the other road user groups

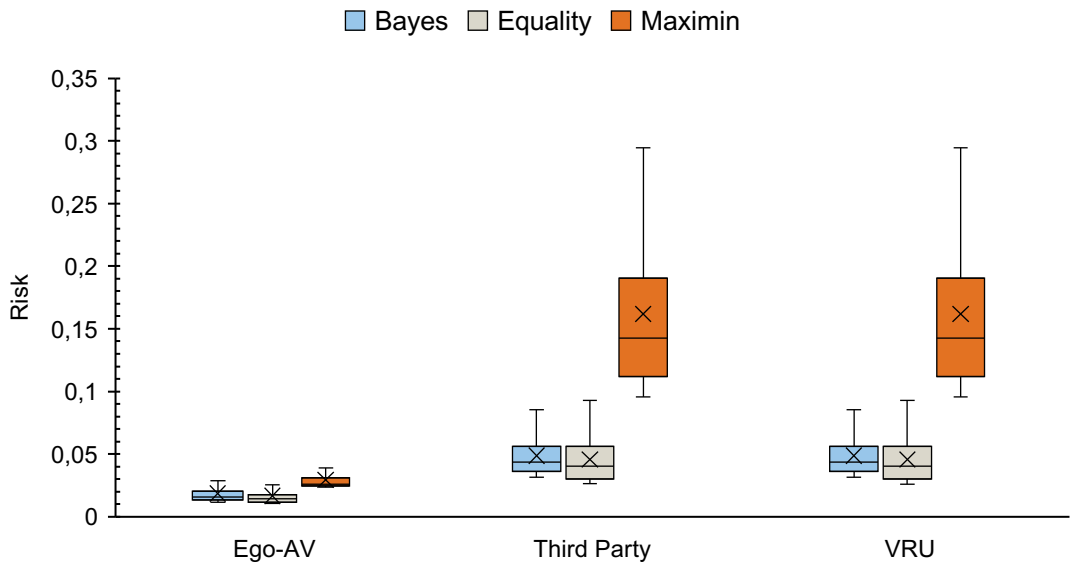


Figure 5.6: Risk distribution of the highest 100 occurring risks. Three different algorithms (Bayes, equality, and maximin) are compared on three different groups of road users (ego-AV, third party, and VRUs). The data are presented as median (Q2) between Q1 and Q3 with horizontal lines. The mean is indicated with a cross, and the whiskers conform to the Tukey style.

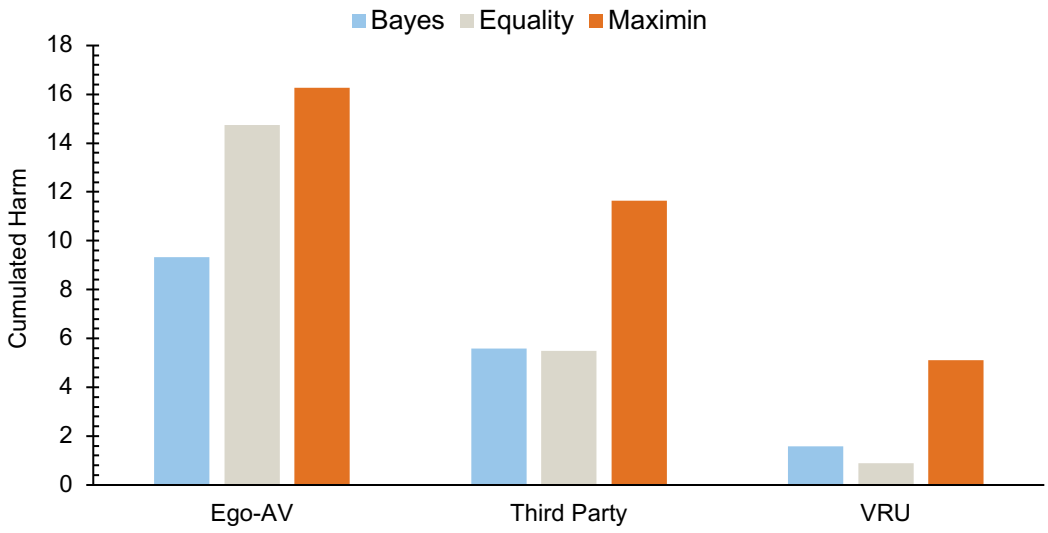


Figure 5.7: Cumulated harm for three algorithms (Bayes, equality, and maximin) as a result of 2,000 simulated scenarios categorized in three different groups of road users (ego-AV, third-party, VRU)

would have indicated. In particular, the observed harm for all three algorithms is highest for the ego-AV. This is essentially due to the fact that the ego-AV can collide not only with other road users but also with static objects or the track boundary. Because of the deterministic behavior of other road users, this is not possible with them. As expected, the Bayes algorithm leads to the lowest overall harm (ego-AV + third party). While the equality algorithm causes a higher amount of harm for the ego-AV compared to the Bayes algorithm, it achieves a similar amount of harm for third-party road users. The effect of the equality principle is particularly evident for the VRUs. Here, it yields significantly lower harm for this group than the Bayes algorithm. Based

on this data, it can be concluded that the equality principle causes a shift in harm from VRUs to motorized road users.

5.2.2 The Ethical Algorithm

In the introduction, the need for an ethical approach to trajectory planning with a fair distribution of risk was motivated in contrast to a selfish risk distribution in trajectory planning. The algorithm as proposed with the parameters of the user survey is denoted as an *ethical* algorithm here since it considers three ethical principles with the purpose of fair risk distribution. A basic algorithm for trajectory planning, as it exists in the state of the art [272], will serve as the basis for the comparison. The same sampling-based planning approach and the same parameterization of the cost function are used. However, the terms of the cost function that explicitly consider risk are not taken into account. This algorithm will be referred to as *baseline*. To mimic a selfish algorithm, another risk distribution principle minimizes the risks of the ego-AV according to Equation 5.2. A corresponding algorithm with $J_{\text{Risk}} = J_S\ (w_B, w_E, w_M = 0)$ is denoted as *selfish* algorithm in the following. Table 5.2 gives an overview of the three variants of planning algorithms with their parameters.

$$J_S(u) = R_{\text{ego}}(u) \tag{5.2}$$

Table 5.2: Overview of algorithm parameters for baseline, ethical, and selfish algorithm

Algorithm	w_B	w_E	w_M	w_S
Baseline	0.0	0.0	0.0	0.0
Ethical	0.53	0.12	0.35	0.0
Selfish	0.0	0.0	0.0	1.0

Similar to the investigation of the Bayes, equality, and maximin algorithm, Figure 5.8 shows the distribution of the 100 highest risk for the ethical, selfish, and baseline algorithms. For all three observed road user groups, the baseline algorithm shows the highest risks. This indicates that considering risks explicitly in trajectory planning is beneficial in terms of risk mitigation. For the ego-AV, the occurring risks with the ethical algorithm are similar, although slightly higher, than with the selfish algorithm. Looking at the overall average risk for the ego-AV shows the difference more clearly: In the case of the ethical algorithm, the average risk R_{avg} is $3.42 \cdot 10^{-5}$, while the selfish algorithm results in $R_{\text{avg}} = 2.29 \cdot 10^{-5}$. The difference between the ethical and selfish algorithms becomes higher when looking at third-party and VRUs. The ethical algorithm leads to the lowest risks in both cases compared to the other two variants.

To complete the analysis, Figure 5.9 shows the cumulated harm as a result of the collisions that occurred during simulation. In line with the risk distributions from Figure 5.8, the baseline algorithm causes the highest harm for all road users, which underlines the effectiveness of considering risks in trajectory planning. Here, as well, using the ethical approach to trajectory planning instead of the selfish approach shifts harm from VRUs to the ego-AV. While the resulting harm for the ego-AV is higher by 0.51 with the ethical algorithm, third-party harm is decreased by 0.84.

As an overview of all discussed algorithms, Table 5.3 provides the resulting harms for Bayes, equality, maximin, ethical, selfish, and baseline algorithms. From an overall perspective, the Bayes algorithm has the lowest overall harm (14.90), followed by the ethical algorithm (15.21) and the selfish one (15.54). In turn, the equality algorithm caused the lowest harm to VRUs (0.89), followed by the ethical algorithm (1.50) and the Bayes (1.58). The idea of the ethical algorithm of shared ethical principles is thus also reflected in the results of the harm distribution, where the ethical algorithm is shown to be a combination of the different algorithms in terms of their effects on trajectory planning.

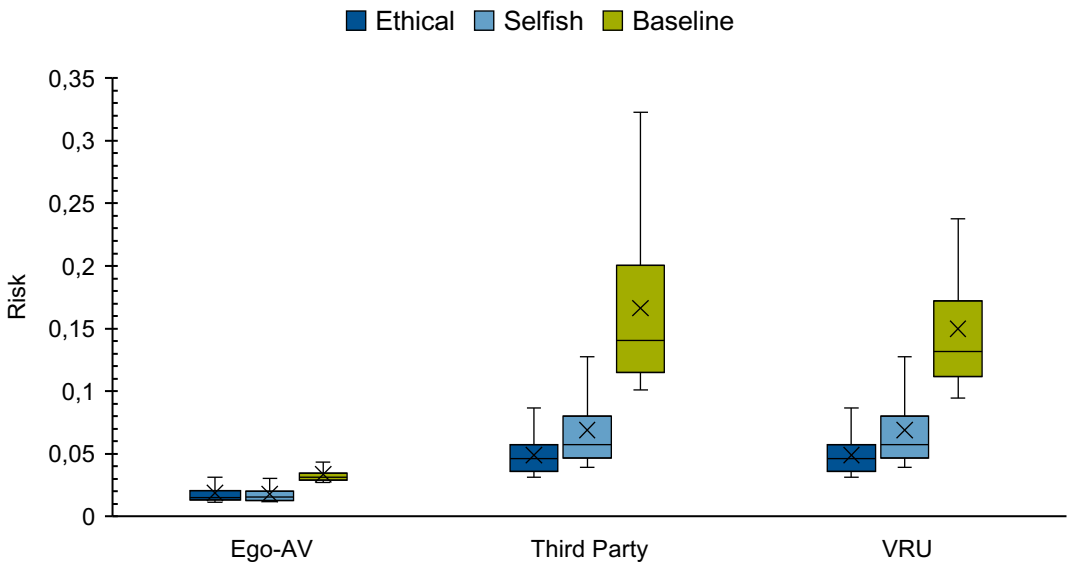


Figure 5.8: Risk distribution of the highest 100 occurring risks. Three different algorithms (ethical, selfish, and baseline) are compared on three different groups of road users (ego-AV, third party, and VRUs). The data are presented as median (Q2) between Q1 and Q3 with horizontal lines. The mean is indicated with a cross, and the whiskers conform to the Tukey style.

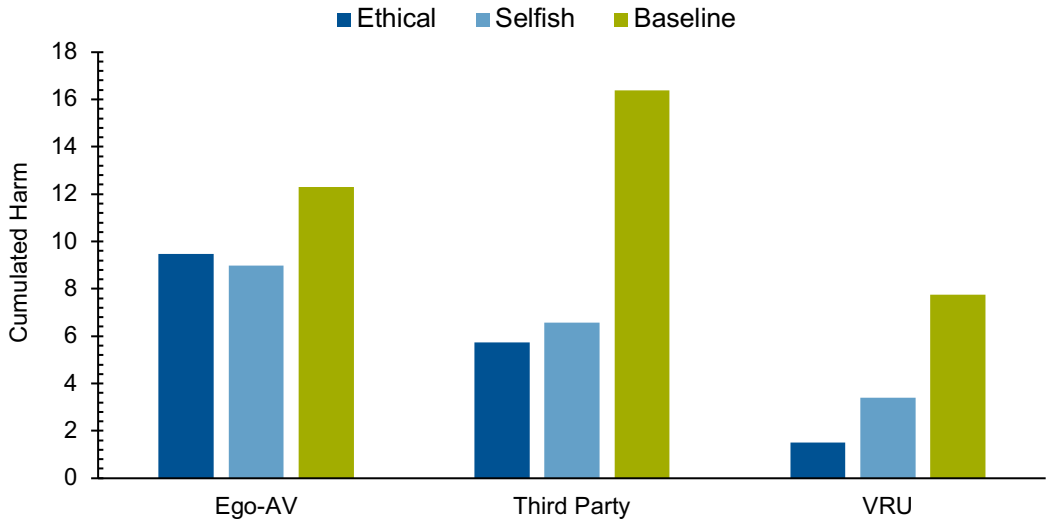


Figure 5.9: Cumulated harm for three algorithms (Ethical, Selfish, and Baseline) as a result of 2,000 simulated scenarios categorized in three different groups of road users (ego-AV, third-party, VRU)

The various configurations of planning algorithms have been shown to lead to different results in terms of risk and harm from an empirical perspective. Comparing the resulting trajectories for these algorithms qualitatively in different scenarios reveals that these empirical differences can hardly be noticed in a scenario. Similar to Figure 1.2, different approaches (e.g., selfish and ethical) only lead to small differences in the planned AV trajectories. However, these seemingly small differences have been shown to lead to significantly different results from an empirical point of view. Therefore, to further explore the differences between these ethical principles, their correlation will be examined in the next section.

Table 5.3: Cummulated harms as a result of simulated accidents on 2,000 scenarios using various algorithms (Bayes, equality, maximin, ethical, selfish, and baseline) for different road users (ego-AV, third-party, VRU, and altogether in total). The lowest harm for each road user group is in bold.

	Bayes	Equality	Maximin	Ethical	Selfish	Baseline
Ego-AV	9.32	14.74	16.26	9.48	**8.97**	12.3
Third-party	5.58	**5.50**	11.65	5.73	6.57	16.38
VRU	1.58	**0.89**	5.10	1.50	3.41	7.75
Total	**14.9**	20.24	27.91	15.21	15.54	28.68

5.2.3 Correlation Analysis

So far, the effects of various principles on risk distribution, as well as the resulting harm, have been investigated in the simulation. The following will examine the proposed cost function terms of all principles in more detail. The objective of the linear cost function is to use linearly independent terms. If various principles or a combination of principles lead to the same decision in trajectory planning, then the proposed cost function terms with ethical considerations do not add novelty to the cost function, and the cost function could be simplified. This is why the correlations of the cost terms below are investigated. Unlike the empirical results presented so far, the correlation analysis is independent of the weighting parameters chosen.

For correlation analysis, the single terms of the cost functions are compared pairwise. The Pearson Correlation Coefficient (PCC) serves as an indicator for the linear correlation of two cost terms, J_1 and J_2 with their respective mean values $\overline{J_1}$ and $\overline{J_2}$. Figure 5.10 shows the pairwise PCCs between all introduced principles. For this purpose, the PCC_S for the set of 2,000 scenarios S_S were determined for all possible combinations of cost terms. As shown in Figure 5.10, the PCC as the mean value of all scenarios is calculated according to Equations 5.3 and 5.4.

$$PCC(J_1, J_2) = \frac{\sum_{s \in S_S} PCC_S(J_{1,s}, J_{2,s})}{|S_S|} \tag{5.3}$$

$$\text{with} \quad PCC_S(J_{1,s}, J_{2,s}) = \frac{\sum_{i=1}^{n} (J_{1,i} - \overline{J_1})(J_{2,i} - \overline{J_2})}{\sqrt{\sum_{i=1}^{n} (J_{1,i} - \overline{J_1})^2 \sum_{i=1}^{n} (J_{2,i} - \overline{J_2})^2}} \tag{5.4}$$

Among the risk distribution principles, high correlation coefficients are found between the Bayes, equality, and selfish principles. As expected, the responsibility principle is negatively correlated to the remaining terms due to the negative sign in the cost function. The terms of costs that are not related to risk, such as the costs of velocity and lateral distance, are largely uncorrelated ($|PCC| < 0.2$) with the other principles. Only the maximin principle and the velocity costs show a stronger correlation since the velocity is an explicit part of the harm model, which largely describes the maximin costs.

Next, it is interesting to see whether the trade-offs of the ethical principles that were theoretically motivated by Figure 3.9 occur in the same way in real AV trajectory planning. Therefore, an investigation is conducted into the variations in trajectory selection based on the principles. This involves an analysis of the distances at the planning timestep $t = 2.0\,s$ of the trajectories that would be chosen by each principle, as illustrated by Figure 5.11.

Figure 5.12 illustrates these distances for the Bayes, equality, and maximin principles. Again, the distance values shown in this figure are mean values for each scenario. It becomes clear that even if there is a high

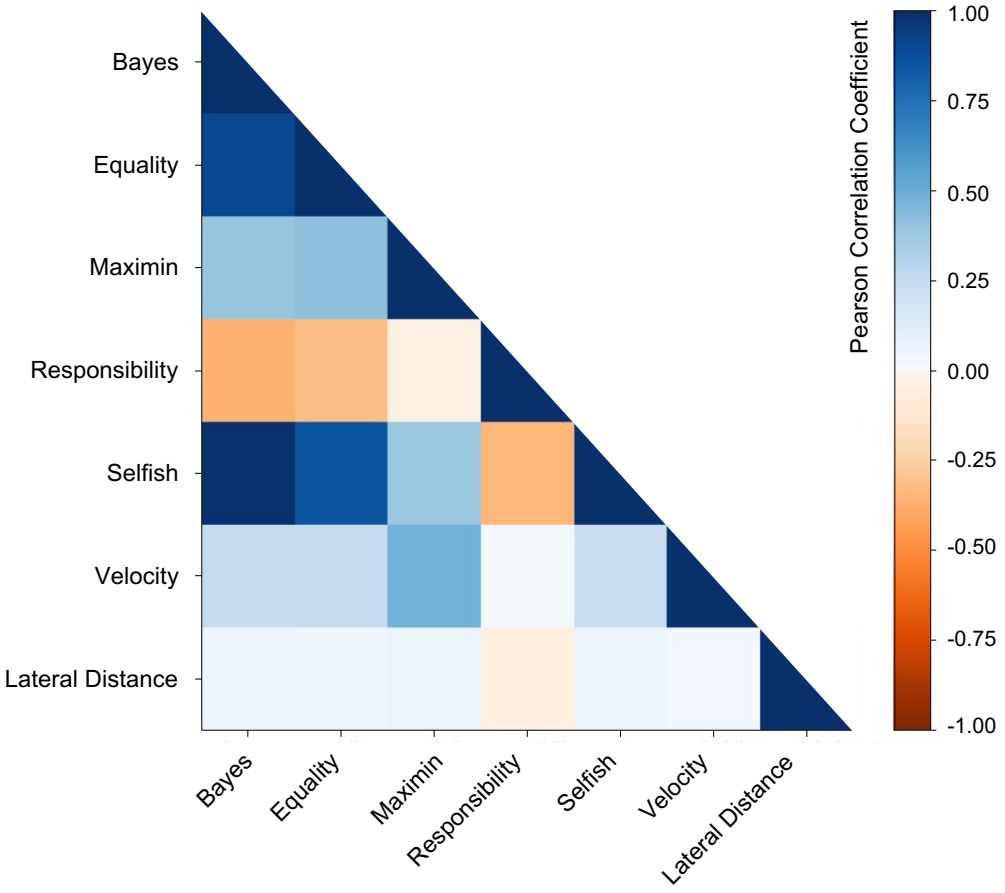


Figure 5.10: Correlation matrix with pairwise Pearson Correlation Coefficient for all cost terms over all 2,000 scenarios.

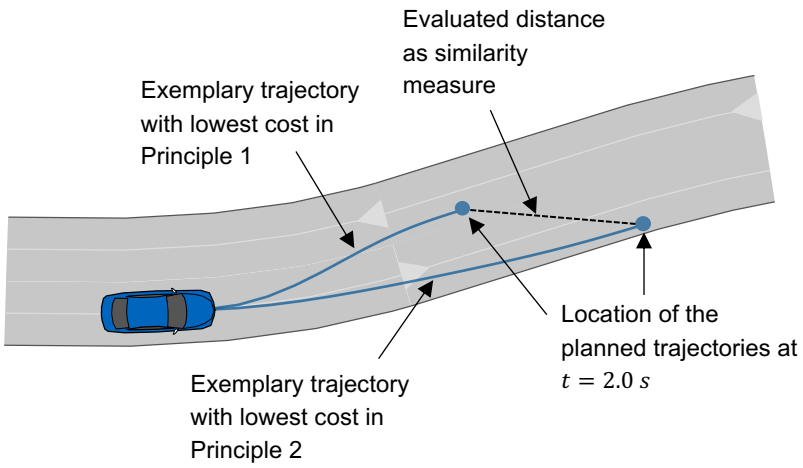


Figure 5.11: To analyze the similarity of various ethical principles, the distances at the planning horizon are evaluated for trajectories that would be preferred by a principle. The resulting distances are shown by Figure 5.12.

degree of correspondence in the trajectory choice, there are scenarios where the deviation of the trajectories is greater than $10\ m$ on average for all principles. These scenarios could be particularly interesting from an ethical point of view, as different principles lead to significantly different choices in the trajectory, and thus, the ethical trade-off arises. This observation will be revisited in the outlook in Section 6.5.

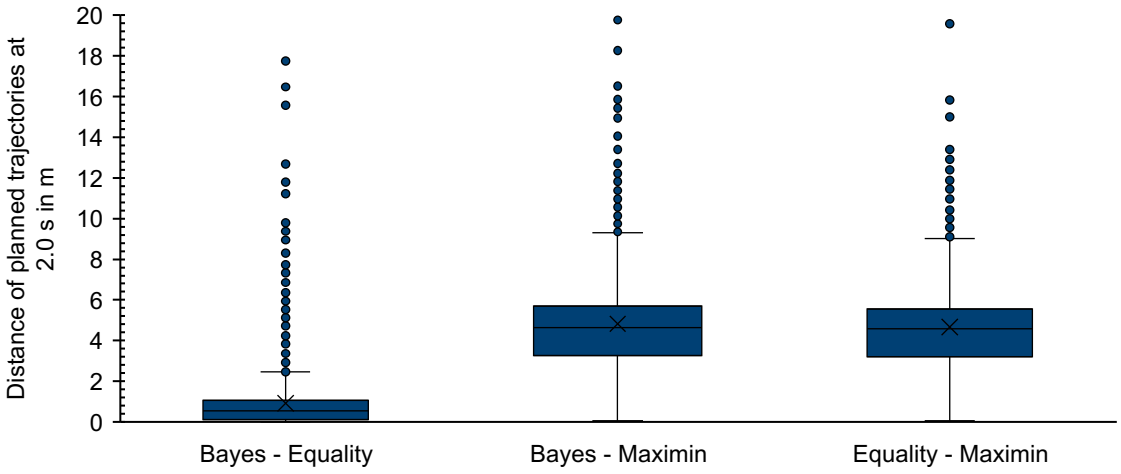


Figure 5.12: A comparison of ethical principles is conducted by evaluating the average deviation of their respective selected trajectories utilizing 2,000 simulated scenarios. While there exist scenarios (data points) with deviations exceeding $15\ m$ for all combinations, the maximin principle exhibits the largest average deviations. The data are presented as median (Q2) between Q1 and Q3 with horizontal lines. The mean is indicated with a cross, and the whiskers conform to the Tukey style. Small circles represent outliers. The figure was adapted from a previous author's publication [268].

These results are underlined by analyzing the multiple correlation coefficients for the ethical principles. So far, the correlation analysis has been limited to a pairwise comparison of principles. To investigate linear dependencies between the proposed ethical principles, possible linear combinations must be considered as well. Therefore, the coefficient of multiple correlations [301] reflects how well a risk distribution principle can be predicted using a linear function of the two remaining principles. Table 5.4 shows each principle's correlation values and underlying linear regression function. The maximin principle is nearly uncorrelated in both directions, with a correlation factor and linear weighting terms lower than 0.1. The Bayes and equality principles show stronger correlations. However, since the maximin principle does not add to the correlation with linear factors smaller than 0.002 in both cases, the previously presented pairwise investigation reflects the correlation between these principles well enough.

Table 5.4: Multi-correlation coefficients for each ethical principle based on their resulting linear regression.

Risk Distribution Principle	Correlation	Linear Regression
Bayes	0.431	$J_B^* = 0.716 J_E + 0.0013 J_M + 1.429$
Equality	0.431	$J_E^* = 0.256 J_B + 0.0007 J_M + 0.238$
Maximin	0.063	$J_M^* = 1.011 J_B + 1.436 J_E + 42.66$

5.3 Maximum Acceptable Risk

Section 3.3 introduced the concept of maximum acceptable risk, with the hypothesis that this approach offers guidance for behavior and decision-making in trajectory planning. This perspective broadens the scope beyond the issue of risk distribution and encompasses the emergence of risk, which directly influences the risks that arise in the context of autonomous driving. Deriving a corresponding behavior of the AV using a maximum acceptable risk has more far-reaching potential than the question of ethical risk management. While touching this potential in terms of the technical implications of the principle, the focus here remains on examining the principle against the background of ethical trajectory planning. To showcase the effect of that principle and refer to the hypothesis of behavior guidance, the next section begins with a qualitative example and then illuminates the effects of that principle in a quantitative context.

5.3.1 Qualitative Example

The fundamental concept of the proposed idea is to move away from prescribing specific behaviors for AVs, e.g., by defining rules and evaluating safety or risk metrics a posteriori. Instead, the approach suggests specifying an acceptable level of risk and deriving the AV's behavior from it. This involves using a risk threshold to determine appropriate actions, such as setting a suitable velocity or deciding whether to perform an overtaking maneuver. To illustrate this, a scenario is employed involving a slow-moving scooter and an AV following the proposed trajectory planning algorithm on a rural road with heavy oncoming traffic (depicted as trucks).

Within this scenario, a comparison is drawn between two distinct configurations: in configuration a) there is no maximum acceptable risk set, while in b) R_{max} is set to 10^{-7}. This value is chosen as an example to showcase the different behaviors. The initial situation is shown on the left side of the figure for both configurations. The colors indicate the cost associated with the sampled trajectories for overtaking or not overtaking the scooter. In configuration a), trajectories that involve overtaking have lower costs (green) due to the significant speed difference between the AV's target speed and the relatively lower velocity to follow the scooter. The cost function, which balances safety and efficiency, favors overtaking here. The AV's willingness to take risks increases the more its current velocity decreases or, in general, deviates from the target velocity. Consequently, the algorithm plans to overtake the scooter at t_1 and t_2 in the scenario without maximum acceptable risks.

However, when the same planning algorithm and parameters are used, but with $R_{\text{max}} = 10^{-7}$ (configuration b), the behavior changes. Trajectories with low costs for the overtaking maneuver exceed the maximum acceptable risk. Consequently, they are classified as invalid, as indicated by the blue color in Figure 5.13. As a result, the AV refrains from performing an overtaking maneuver and maintains a safe distance behind the scooter. This happens only on the basis of risk without any scenario-specific rule (e.g., safety distance) being set that prescribes this behavior. This example highlights the appropriateness of setting a maximum accepted risk for maneuver-level decisions. Nevertheless, supplementary experiments reveal that the duration of the planning horizon plays a crucial role in this context. If the planning horizon is insufficiently short (e.g., $1\,s$ instead of $2\,s$ as in this case), the risks associated with an overtaking maneuver cannot be properly assessed. This may result in the AV entering situations where it could not anticipate the risks in advance.

5.3.2 Empirical Evaluation

The concept of maximum accepted risk, as illustrated in the previous example, has implications for behavior in various scenarios that impact the overall results in terms of risk allocation. This can be made visible by means of an empirical evaluation. To assess these effects, a simulation evaluates the proposed trajectory

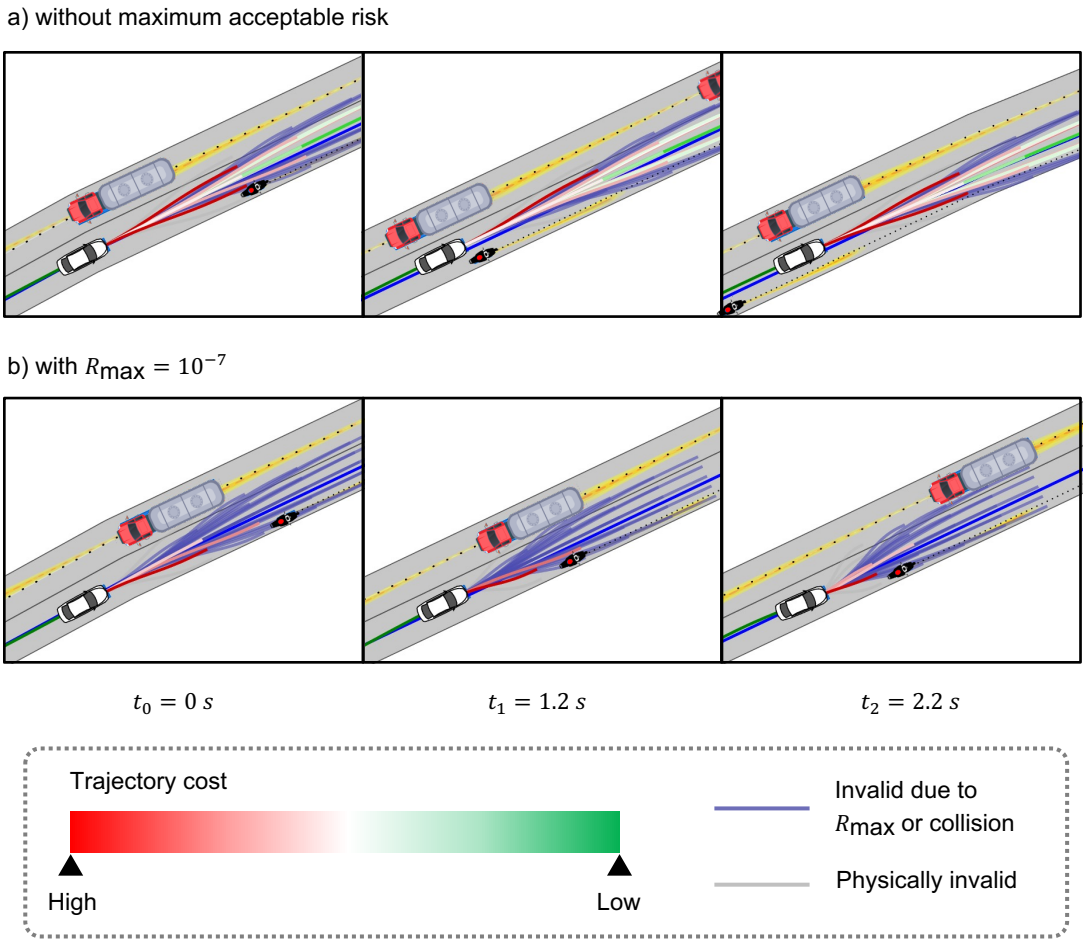


Figure 5.13: An overtaking maneuver serves as a qualitative illustration of the impact of a maximum acceptable risk in trajectory planning. The maneuver is executed using two algorithmic configurations observed at different time steps: In the absence of a maximum acceptable risk constraint (a), the white AV overtakes the slower scooter. However, when $R_{\text{max}} = 10^{-7}$ is introduced to the same algorithm, the overtaking maneuver is deemed unsafe, forcing the AV to remain behind the scooter (b). The figure was adapted from a previous author's publication [270]

planning algorithm on 2,000 scenarios. A comparison was conducted across various values for R_{max}, where the resulting risk distribution served as a metric for evaluation. This analysis considered the risks faced by all road users collectively at each simulation timestep, taking into account the specific configuration. In order to also shed light on the effect of maximum risk as a standalone principle, the focus is first on the baseline planning algorithm without the risk cost function. In a second step, the ethical planner as in the previous Section 5.2.2 is then extended to include the concept of maximum accepted risk, and, in particular, the implications with regard to risk emergence and distribution are considered.

Figure 5.14 presents the tail distribution of the calculated risks for different values of R_{max} building up on the baseline planning algorithm. The tail risk distribution shows the percentage of planning timesteps at which the risk is greater or equal to a certain value (horizontal axis). Consequently, at higher risks, lower values in the distribution are aimed for. As R_{max} decreases incrementally from ∞ (which equals no maximum acceptable risk) to 10^{-6}, distinct discontinuities emerge at the corresponding values of R_{max}. As shown by Figure 5.14 all curves show a discontinuity at their respective value of R_{max}. This observation is reasonable because decision-making based on the maximum acceptable risk is not continuous but rather exhibits a discrete nature.

A further part of the investigation is how R_{max} impacts the actual risks. Hence, the observation centers on R_{avg}, which denotes the average risk across all timesteps within the 2,000 scenarios. Ideally, these two values are correlated, and reducing R_{max} results in lower average risk in trajectory planning. However, looking at the comparison of R_{max} and R_{avg} in Figure 5.14 based on the simulation reveals that lowering R_{max} does not necessarily lead to lower average risk. Comparison between $R_{\text{max}} = 10^{-4}$ (orange line) and $R_{\text{max}} = 10^{-5}$ (black line), for example, shows that, although the occurrence of risks higher than 10^{-5} is less frequent with $R_{\text{max}} = 10^{-5}$ (3 % versus 4 %), the average risk, in that case, is higher than with $R_{\text{max}} = 10^{-4}$ ($7.25 \cdot 10^{-5}$ versus $7.13 \cdot 10^{-5}$). This is due to the fact that higher risks (10^{-4}) are more likely to occur in this specific case compared to when $R_{\text{max}} = 10^{-4}$. A further characteristic that indicates this observation is the intersections of the curves for different R_{max} values. If there is a sufficiently strong correlation between R_{max} and the risks that actually arise, such intersections should not occur.

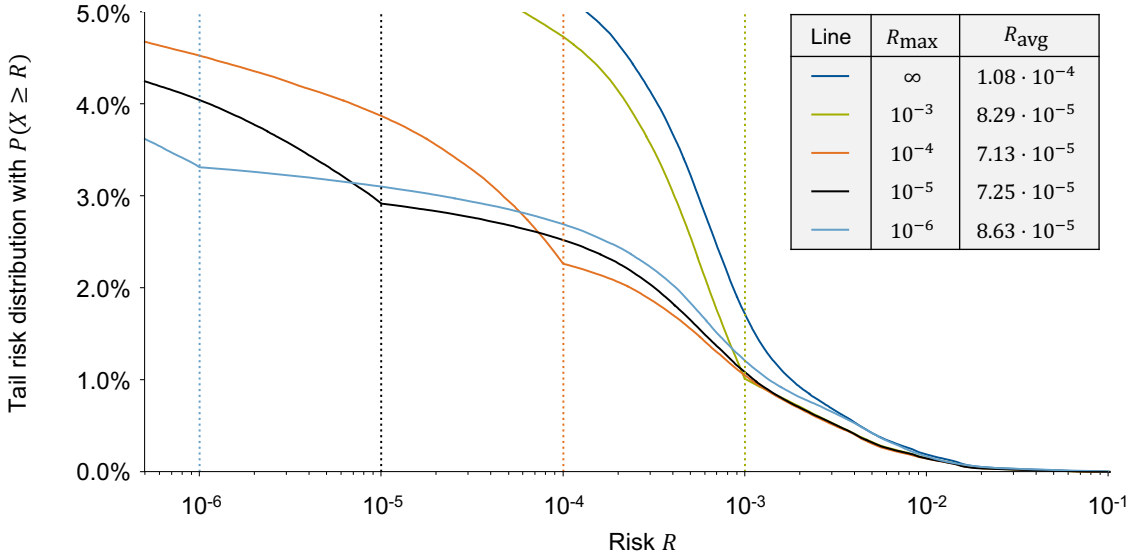


Figure 5.14: Tail distribution of actual risks occurring during simulation with 2,000 CommonRoad scenarios with various values for maximum acceptable risks in combination with the baseline planning algorithm. The corresponding values for R_{max} are shown as vertical lines with the same color.

In the next step, the maximum acceptable risk is investigated along with the ethical risk distribution function from Equation 3.11 with the risk distribution parameters from Table 5.2. Thus, the ethical algorithm proposed in Chapter 3 is evaluated in the following with all proposed principles in combination. Figure 5.15 shows the tail risk distribution of this algorithm in comparison to the previous configuration with only the maximum acceptable risk. It is noticeable that, on average, the risks are lower by a factor of 3 to 4 than before. Where intersections within the distribution curves have previously occurred, it can be seen that the curves at the points for R_{max} are close to the curve of the lowest R_{max} value. The risk proportion values for risks greater than R_{max} are also close to each other. This is due to the fact that in the case of R_{max} being exceeded, the cost function does not consider a trade-off between risk and mobility anymore but only focuses on risk. The different states that the AV is in due to previous behavior can explain the slight discrepancies in the risk distribution.

Even though for $R_{\text{max}} = 10^{-6}$, the average risk is slightly higher than for $R_{\text{max}} = 10^{-5}$, a clearer correlation between R_{max} and R_{avg} can be seen.

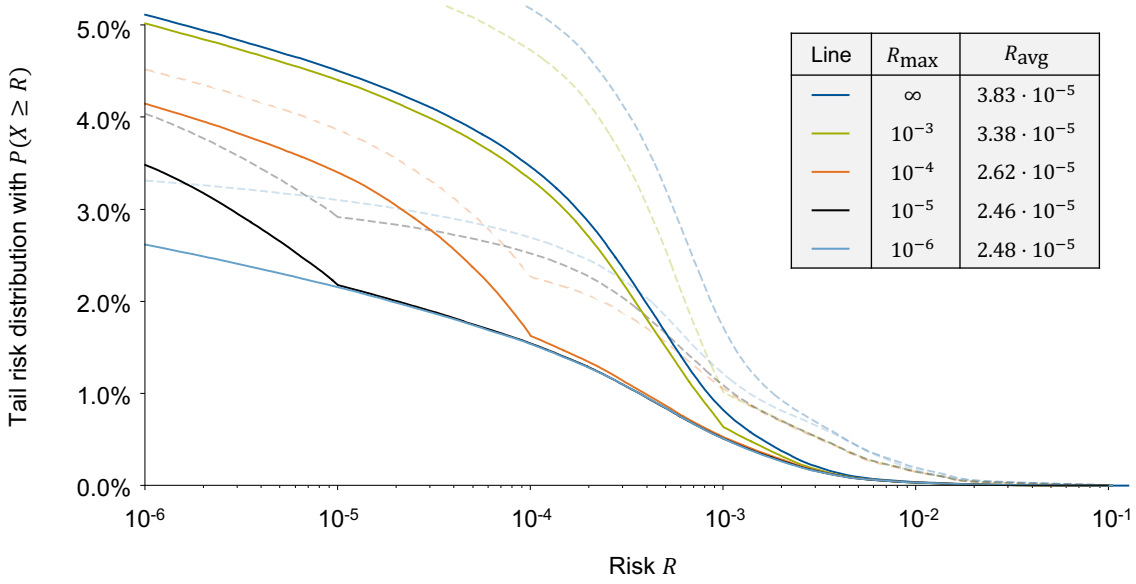


Figure 5.15: Tail distribution of actual risks occurring during simulation with 2,000 CommonRoad scenarios with various values for maximum acceptable risks in combination with the ethical planning algorithm. The results when using the baseline algorithm from Figure 5.14 are shown as transparent dashed lines for comparison.

Figure 5.16 evaluates the corresponding cumulated harms as a result of the simulation runs with various values for R_{max}. The lowest value for harm is achieved with $R_{\text{max}} = 10^{-5}$. This shows an improvement of 3.21 in harm compared to the initial situation without R_{max} ($H = 15.21$). Naturally, the number and severity of accidents should decrease with decreasing R_{max}, which should then be directly reflected in the cumulative harm. With lower accepted risk, the vehicle would move more slowly in its surroundings or, in extreme cases, not at all, in order not to take any unaccepted risks. However, this is not possible in the simulation because the scenarios and the behavior of the other road users are deterministic. Driving slowly or stopping could eventually lead to more accidents under certain circumstances. This also explains the increase in average risk and cumulative harm when reducing R_{max} from 10^{-5} to 10^{-6}. The average velocities, which even increase slightly between $R_{\text{max}} = 10^{-3}$ ($8.68\ m/s$) and $R_{\text{max}} = 10^{-6}$ ($8.75\ m/s$), underline this hypothesis.

The distribution of harm between various road user groups remains largely unaffected, as Figure 5.16 shows. The maximum accepted risk takes into account the risks of all road users and does not explicitly impact the risk distribution. Accordingly, the distribution of risk is significantly influenced by the risk cost function with the corresponding distribution principles.

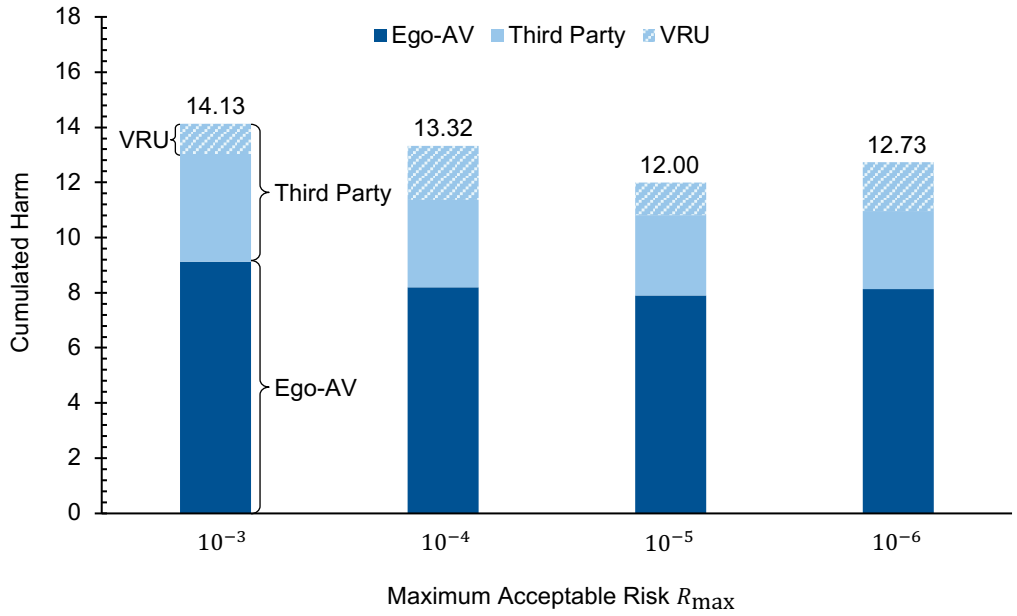


Figure 5.16: Cumulated harm as a result of 2,000 simulation scenarios with various values for $R_{\max}$. The colors indicate the distribution of harm to the associated road user groups ego-AV, third-party road users, and VRUs.

6 Discussion

The discussion should serve to critically question the methods and results presented up to here. To this end, Section 6.1 will first take up the research questions formulated as a result of Chapter 2 and briefly summarize and discuss the answers proposed in the course of this thesis. The following Section 6.2 will then discuss the validity of the proposed ethical framework under various aspects, such as the proposed risk model or the selected guiding ethical principles. Section 6.3 refers back to Section 2.2.4 that formulated requirements based on the legislation and ethical standards. This section will examine the extent to which these requirements have been met in this work. A further important requirement next to the legislation was the applicability in a real-world application, which will be investigated in Section 6.4. Finally, with recommendations for industry and policymakers, Section 6.5 will propose how to proceed with this work as a basis for ensuring ethical behavior of AVs in the future.

6.1 Review of Research Questions

Section 2.4 prompted the research question of how an ethical algorithm with fair risk assessment can be realized for the trajectory planning of AVs. This overall research question was broken down into five subquestions, which, in summary, provide the answer to the guiding research question. Therefore, the following section will refer back to the five questions and present and discuss their proposed answer here.

6.1.1 Ethical Algorithm Based on Legal Requirements

Q1: *How can ethical aspects be considered in algorithms for AV decision-making based on legal requirements?*

To answer the first research question, the legal and ethical foundations were explained in Chapter 2. This resulted in a new problem statement than what was mainly worked on before: instead of solving or discussing abstracted dilemma situations, such as the trolly problem, the whole complexity of AVs must be considered. The discussion of life and death decisions, possibly even on the basis of discriminatory factors, must be overcome since it does not align with current legislation nor reflect the technical aspects of AVs. This incorporates algorithmic and technical limitations, as well as handling risks and uncertainties. Consequently, in a comprehensive analysis of ethical fundamentals in Section 2.2.2, the application of risk ethics was evaluated to be promising for decision-making in AVs. For ethical decision-making, ethical principles from the context of risk ethics should, therefore, be integrated into a trajectory planning algorithm for AVs. To achieve this, the foundations must be laid on the technical side so that decision-making in trajectory planning can be based on uncertainties. Hence, the research question Q1 has been sufficiently answered.

6.1.2 Ethical Principles for AV Decision-making

Q2: *Which ethical principles may be considered in the trajectory planning of autonomous vehicles?*

With the application of risk ethics, the ethical challenge has shifted to the question of fair risk distribution in the trajectory planning of AVs. This raises a distribution problem for which parallels can be found in the literature, such as in organ donation or in connection with radiation. Section 3.4 showed in line with the recommendation of the EU expert group [11] that a single principle for risk distribution is not sufficient. Instead, a combination of shared ethical principles is demanded. A structured analysis of existing principles in Section 3.4 against the background of a) technical feasibility and b) the aspect of fairness (Q4) has shown the following ethical principles to be reasonable:

- Maximum acceptable risk (Section 3.3)
- Bayes principle (Utilitarian principle) (Section 3.4.1)
- Equality principle (Section 3.4.2)
- Maximin principle (Section 3.4.3)
- Responsibility principle (Section 3.4.4)

Further discussion of the selected principles will be part of Section 6.2.2 in the discussion of the framework's validity.

6.1.3 Implementation of Ethical Principles

Q3: *How can an ethical concept from several ethical principles be implemented in an algorithm for AV trajectory planning?*

To guide AV decision-making by the presented principles requires a definition and quantification of risk first. Analysis of the legislation landscape revealed that taking into account the vulnerability of road users is an important aspect here. Therefore, consideration of the two dimensions of collision probability and severity of collisions (personal harm) is crucial. For both variables, corresponding models have been developed according to legal regulations and ethical guidelines (Section 3.2). However, in the case of the probability of collision, only uncertainties in the prediction of the trajectory of other road users have been considered to date. The consideration of the entire propagation of uncertainties through the AV functionalities is an important component but was not in the scope of this thesis. Based on the quantified risks, the derived principles can be implemented in a sampling-based approach to trajectory planning. The basis for this is the mathematical formulation of the principles, which allows either a) to formulate cost terms or b) to derive validity criteria as a form of soft constraints. Hence, the research question Q3 has been answered, but further open points have been identified with respect to other sources of uncertainty.

6.1.4 Fairness

Q4: *How can the aspect of fairness be taken into account in the decision-making process of AVs?*

The selection and implementation of various ethical principles that are technically feasible for the application of AV trajectory planning do not yet consider the question of fairness. In fact, decision-making in the proposed algorithm is dependent on the individual weighting parameters for the respective principles. To proceed in the question of fairness, Rawls' Theory of Justice and the Veil of Ignorance were applied. As a result, a user survey, in which participants do not know their own position, was designed and conducted as an online experiment (Chapter 4). The survey was developed against the background that parameters for the

weighting factors of the principles can be calculated by answering the questions. The survey results with 247 participants suggest prioritizing the Bayes principle, while the equality principle seems to be less important to the participants. Even if the results of the study are not representative because the target group is biased, for example, the concept of fair distribution of risk has been implemented accordingly, and the path to proceed has been shown. Consequently, the work has shown one way to consider the aspect of fairness in the distribution of risk. However, there might be additional theories and approaches to this problem that can be applied here.

6.1.5 Empirical Effects

Q5: *What are the actual effects of using an ethically motivated algorithm for trajectory planning compared to state of the art?*

So far, the state of the art in ethical AV decision-making consists largely of theoretical considerations that are not technically sound and have not been empirically tested for their real-world consequences. The trajectory planning algorithm as proposed in Chapter 3 with the parameters that were the result of the user survey from Chapter 4 was, therefore, evaluated in simulation on 2,000 scenarios. The results, as reported in Chapter 5, show that the explicit consideration of risk in the planning of the AV trajectory is beneficial for all road users in terms of risk exposure. They further indicate that using the proposed ethical approach instead of a selfish risk distribution principle indeed leads to better protection of VRUs. Risks are shifted from VRUs to the ego-AV, whereas the proportion of risk reduction for VRUs is significantly higher than the growth of risk for the ego-AV. Especially the equality principle seems to contribute to this effect, which is desired by the legislation and the EU expert group. Adding a maximum acceptable risk in addition to the cost function that regulates risk distribution affects this distribution only slightly but is suitable to guide trajectory planning from a maneuver perspective and reduce risks.

6.2 Validity of the Ethical Framework

Ethical theories tend to seek ultimate solutions and are intended to be internally consistent [302]. This distinguishes the work of social scientists and philosophers from that of engineers and developers in research, who strive for improvement over the state of the art. This interdisciplinary work was accomplished according to the working principle of engineering: The presented solution does not claim to be a perfect and all-encompassing solution in every case. It is more considered a first suggestion, which should open up new discussions in this field.

The objective of this work was to improve the state of the art. Nevertheless, the shortcomings, both solvable in principle and seemingly unavoidable, are to be revealed in this section. Therefore, the following section will first discuss the underlying risk model (Section 6.2.1) and then the selection of guiding ethical principles (Section 6.2.2) as the novel core components of the proposed algorithm. Another, so far unmentioned shortcoming is the information asymmetry and incompleteness (Section 6.2.3), since information is only available from the AV's point of view, which complicates a moral assessment of risk. Finally, limitations of the Ethical Vehicle Experiment, presented in Chapter 4, will be discussed in Section 6.2.4.

6.2.1 Risk Model

Chapter 2 showed various state-of-the-art risk metrics for trajectory planning, such as the TTC or TTR. The risk model used in this work differs substantially from these metrics in two ways: First, TTC or TTR are independent of the algorithms that provide the inputs for trajectory planning. The risk model, instead,

actively considers uncertainty from previous functions, such as trajectory prediction. On the one hand, this requires extensions in the algorithms to quantify the uncertainties. On the other hand, the non-negligible dependence of the risk on the algorithmic functionality is also reflected. Second, due to the definition of risk as a two-dimensional measure consisting of a collision probability and estimated harm, the vulnerability of various road users can be taken into account. Previous risk measures do not actively consider the vulnerability of road users, which, however, is an essential requirement formulated by the European expert group [11].

While the proposed risk model offers significant advantages, there are also some drawbacks to consider. So far, the model only takes into account prediction uncertainties, leaving uncertainties from other functionalities, such as localization or detection, as open points. Modeling the dependencies of various types of uncertainty can be challenging, especially considering that most perception tasks rely on neural networks, which often have black-box characteristics. Investigating the interplay between dependent and independent uncertainties becomes complex due to this lack of transparency. Additionally, similar to the prediction model discussed in Section 3.2.1, the uncertainties in the proposed risk model are related to the data on which the neural network was trained on. Therefore, selecting the training data becomes crucial to ensure valid coverage of relevant scenarios for the intended application. It is important to note that unseen or underrepresented situations can lead to higher uncertainties in the model outputs. This aspect could be beneficial in addressing the issue of bias towards underrepresented groups, which was briefly mentioned as an ethical concern in Chapter 2.

There are additional manageable shortcomings associated with the prediction and harm model used in this work that warrant attention. The prediction model takes the road network and driveable space as input and generates probabilistic trajectories as output. However, one limitation is that these probabilistic trajectories are not constrained to the driveable space. The model does not explicitly enforce that the generated trajectories adhere to the boundaries of the road or stay within the regions where a vehicle can safely maneuver. As a result, there is a possibility of generating trajectories that extend beyond the boundaries of the driveable space. To model uncertainties in the prediction, the model uses Gaussian distributions. These Gaussian distributions are not inherently constrained to the driveable space. This means that even in areas where a collision with an obstacle is highly unlikely or impossible in reality, there is still a small, albeit theoretically existing, collision probability assigned by the model. The issue arises from the fact that the sum of all existing probabilities in the driveable space does not add up to 1, leading to an underestimation of the collision probability.

To address this limitation, it would be beneficial to explore additional probability distributions that are suitable to constrain the existence probability of trajectories within the driveable space. By utilizing probability distributions that explicitly enforce the constraints of the road boundaries and the safe regions for vehicle movement, it is possible to improve the accuracy and reliability of the collision probability estimation. The introduction of such constrained probability distributions could be a promising avenue for enhancing the prediction and harm model in this work. Ensuring that the generated trajectories align more closely with the actual driveable space would lead to more accurate collision probability estimations and a better understanding of the associated risks in different driving scenarios.

6.2.2 Selection of Guiding Ethical Principles

In the literature, various distribution principles exist. While some of them can be adapted to address the problem of risk distribution in autonomous driving, others are not appropriate here. One well-known example of a distribution problem is the allocation of resources, such as money. Examining these established distribution principles can provide insights that can be applied to the current application of risk distribution. In the context of resource distribution, it is common for the total amount of money to remain constant while being distributed among multiple parties. However, when it comes to risk distribution, there is a fundamental difference. The

total amount of risk can vary across different options or scenarios. This variability in the total risk motivates the utilization of principles such as the Bayes principle, which explicitly accounts for the varying total amounts of risk. Considering the principles employed in resource distribution can identify parallels and draw useful insights for risk distribution. Some principles may be directly applicable to the risk distribution problem at hand, allowing for their adaptation and utilization. These principles provide a framework for effectively allocating and distributing risk among different options or scenarios. However, it is important to note that not all distribution principles from the resource allocation domain may directly align with the requirements of risk distribution. In such cases, further principles specific to risk distribution may need to be identified or developed to address the unique characteristics and challenges associated with distributing risk.

Section 3.4 presented the ethical principles that were explicitly considered as part of the trajectory planning algorithm. However, there are additional principles that could be applicable to this particular application of trajectory planning and risk distribution. Some principles may have been considered implicitly in the development of the algorithm. Table 6.1 gives an overview of distribution principles as the outcome of extensive literature research. For further information, the reader is referred to the author's publication [303]. While Chapter 3 presented and justified the ethical principles that have been integrated into the algorithm, the following will discuss the remaining principles that have not been considered.

Table 6.1: Overview of distribution principles from the literature that can be applied to the task of risk distribution.

Principle	Description	Consideration in ethical framework
Altruism	Minimize only the risk of others, even if own risk increases	No
Apriori consent	Every stakeholder express their consent in advance	In parts
Egoism	Minimize only own AV-risk, even if other risks increase	No
Equality	Equal distribution of risk	Yes
Maximin	Minimize greatest possible harm no matter how likely it is	Yes
Randomness	Choose randomly between multiple options	No
Responsibility	Those who caused the risk should carry the risk	Yes
Time	Short-term effects vs. long-term effects	No
Thresholds	Limitation of the risk by a threshold value that should not be exceeded	Yes
Utilitarianism	Minimization of the overall risk	Yes

Altruism & Egoism

Altruism or Egoism as distribution principles have both not been considered. Not considering egoism for risk distribution is one of the main motivations of this work and is straightforward. In Section 5.2, the selfish principle was introduced to reflect egoism as a contrasting algorithm for comparison with the ethical approach. Both principles, egoism and altruism, require the knowledge of the ego position within a scenario. This contradicts the applied approach to fairness, according to Rawls, which is why the altruistic principle was not considered as well.

Apriori Consent

Obtaining the consent of each stakeholder in advance, as required by the principle of apriori consent, is not feasible for each decision the AV has to make. The trajectory planning step is performed ten times per second so that no communication method format would meet this time requirement. The approach to fairness, which is implemented by means of the study as presented by Chapter 4 follows a similar idea as apriori consent. Therefore, the experiment can be extended to explicitly asking for consent. However, this consent is not for

each decision but is represented by the parameters guiding the decision. Thus, consent could be achieved on the mechanisms and principles that guide the AV decisions but not on the AV decisions directly.

Randomness

The principle of randomness can be applied as a distribution principle in various contexts, including risk distribution. When employing the principle of randomness, the allocation of risks is determined through a random selection or chance-based mechanism. Randomness is often considered a fair and unbiased method [54] that could add equality. The lack of transparency in favor of opacity can be seen as beneficial to avoid human judgement [304]. Furthermore, it is easy and clear to implement and would not add remarkable calculation time. However, the objective of trajectory planning is to find optimal trajectories according to a cost function or similar mechanisms. Consequently, a seemingly best trajectory can be determined from the AV's point of view. Adding randomness could then end in not executing the apparent best trajectory, which might be even more difficult to justify. So, the argument of opacity does not hold when it comes to trajectory planning. The equality aspect of including randomness is explicitly considered by the equality principle while not considering discriminatory features at the same time. For this reason, the principle of randomness is neglected here.

Time

The idea of considering time as a risk distribution principle is to prioritize immediate risk over risk that is further in the future. This seems reasonable since the trajectory is recalculated every $0.1\ s$, and thus, a calculated trajectory is executed only for a short time period. This approach is not new in trajectory planning and is suitable for all costs and not only for those who are connected to risk. However, this adds another hyperparameter and requires knowledge of the risk over time. The presented approach does calculate the risk over time, but Section 3.2 showed that for dependent risks, the maximum risk is a good representative of a trajectory. Using a timing discount factor would couple risk to other cost terms along the time dimension, preventing separate consideration of risk using the maximum operator. For this reason, as this principle would complicate the evaluation of the proposed risk distribution and does not guide the distribution itself, this principle was neglected in this work.

Completeness of Selected Principles

Finally, the question arises to what extent these selected principles can be considered as complete. The presented method, which is based on a selection of principles that were found in the literature, is not suitable to guarantee completeness here. However, an indicator of proper model accuracy could be given by the compliance analysis as part of the Ethical Vehicle Experiment, as in Chapter 4. If all user decisions at a large number of questions can be mapped using the model with the proposed distribution principles, this could indicate if the selected principles provide a valid ground for ethical assessment. Section 5.1 showcased this method and indicated that the moral views of the majority of survey participants could be modeled using the selected principles.

However, there are aspects that were neglected in the study, such as the consideration of the temporal domain in decision-making under risk. Using the time principle, however, was discussed to be reasonable but brings technical difficulties. Ultimately, the compatibility with the trajectory planning algorithm, in which the ethical principles make up only a part of the planning objective, is an important constraint to consider when evaluating the aspect of completeness from an ethical point of view.

6.2.3 Information Asymmetry and Incompleteness

Ethical decision-making ideally requires various information that impacts the decision. However, in reality, only limited information is available, namely from the AV point of view. In road traffic, every AV has only limited information provided by its sensors. As a result, each AV perceives a scenario with different information, which can be considered as information asymmetry. For example, in Figure 6.1, given a certain sensor range, only the orange vehicle perceives all other vehicles in that fictive scenario. In contrast, each of the others only perceives a fraction of road users. Subsequently, the risk assessment is biased by this field of view, as the risk of other road users cannot be perceived in the same way as the risk of the AV. This raises some implications for ethical considerations.

From the AV point of view, the behavior of Vehicle 2 (V2) in Figure 6.1 cannot be determined as a function of the presence or behavior of Vehicle 3 (V3) since the AV does not perceive V3. However, V3 may be relevant to a complete ethical consideration. For example, V2 could thus be perceived as responsible according to the responsibility principle, even though it only responds appropriately to V3. In contrast, the interaction between V1 and V2 can be modeled and considered. The prediction model from Section 3.2.1 considers first-order interactions. For the prediction of probability-based trajectories, the states of all perceived road users are taken into account. More sophisticated interactions (second order or higher) require iteration under game-theoretic aspects. Due to computational time constraints, higher-degree interactions are not covered in this work. However, since road traffic interactions can often play an important role in critical scenarios, a detailed consideration of this issue is recommended for the future.

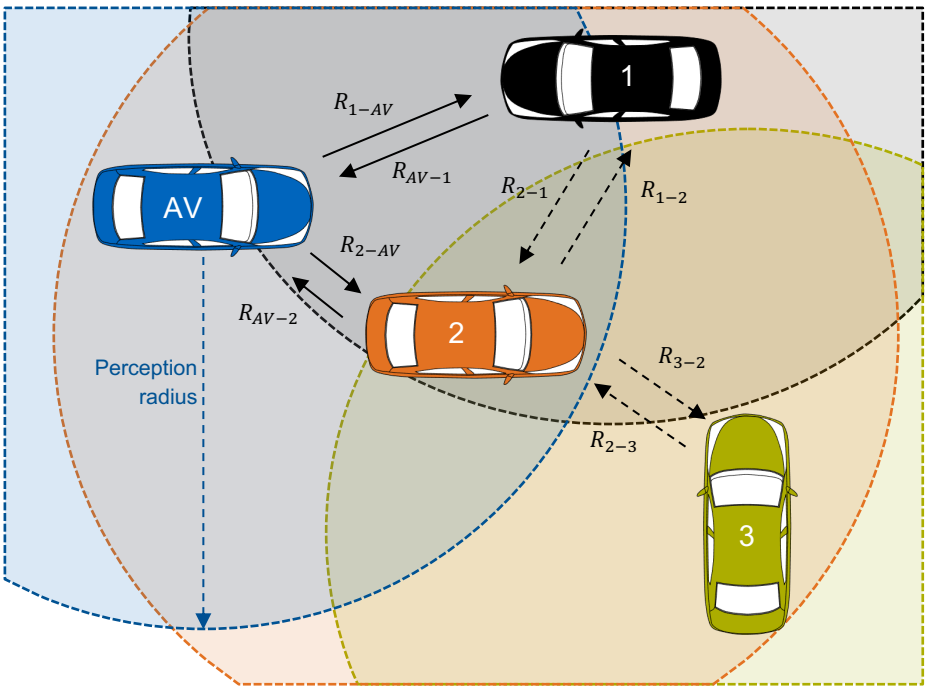


Figure 6.1: Illustration of the limited available information of each vehicle. Due to the limited sensor range, not all relevant road users can necessarily be perceived. The risks indicated with the dashed arrows are not considered in the AV trajectory planning.

The information asymmetry is particularly relevant to the equality principle, which strives to distribute risks equally among road users. Because the AV calculates collision risks with multiple road users while each road user is only assigned a risk for a collision with the AV, the perceived risk of the AV would be systematically higher than the risk calculated for other road users. So, applying the equality principle would fail its purpose due to this bias. Instead, the equality principle is not applied to cumulated risks for the AV but to the single risk terms. So, for the given example, the set of risks S_R to be considered is $\{R_{1-AV}, R_{AV-1}, R_{2-AV}, R_{AV-2}\}$

and not $\{R_{AV}, R_1, R_2\}$. Although the AV perceives both vehicles 1 and 2, the risks R_{1-2} and R_{2-1} that arise from a collision between them are not considered. Since only first-order interactions are considered, these risks are independent of the choice of trajectory and thus do not provide any insight for trajectory planning.

6.2.4 Ethical Vehicle Experiment

The Ethical Vehicle Experiment was developed to provide first insights into how participants would decide in situations where risk must be distributed among road users. Although providing valuable information about this, the study has limitations that should be discussed here. The study was conducted open-access on a nonrepresentative group of people, which led to a bias in the results. However, the objective of the study was to establish a method to connect the moral views of people with algorithmic parameters. Followingly, the focus here will be placed on the study as a method of gaining the desired parameters and less on the execution or the results.

The experiment investigated three principles as dimensions by asking nine questions to the participants. The core of the presented approach is that the model for generating and evaluating the questions corresponds to the model of ethical principles in trajectory planning. This ensures that there is no explicit bias towards one of the principles when using the parameters in the trajectory planning algorithm. As the proposed method uses binary questions, the questions provide a discretization to the solution space. The more questions asked, the smaller the discretization of the resulting parameters (Figures 4.7 and 4.8). The experiment, as conducted, leaves 512 (2^9) unique combinations to answer the questions, which serves as an indicator of the discretization. Thus, the more questions are asked, the more precisely the parameters can be determined as a result. However, this also means that the experiment takes more time, which could decrease the willingness of people to participate without further incentives. Another factor to be considered here is the requirement for risk values that are as simple as possible, such as only one decimal number after the decimal point, as in the experiment conducted. This limits the number of possible questions and the discretization accordingly.

The study was carried out as an investigation of three distribution principles. However, more principles might be necessary for an overall ethical assessment. Adding more principles as additional dimensions increases the number of necessary questions exponentially if the same discretization should be achieved. With more than three principles, this presents challenges. It might, therefore, be worthwhile to investigate the relationships between the principles and, if appropriate, to consider them in a decoupled way. The responsibility principle could add even more complexity. As presented in Section 3.4.4, there are various dimensions of responsibility, e.g., different kinds or severities of rule breaks. This requires additional weighting, which would also open up a new dimension. One way to counter this challenge, in addition to considering the principles in a decoupled manner, is to integrate multiple gradual response options. Figure 6.2 gives an exemplary question that focuses on the trade-off between the responsibility principle on one hand and the Bayes and equality principle on the other hand. Since there is no separation of risks into collision probability and harm, the maximin principle is neglected here. Instead of two options, the example provides five options from A to E. The next step in this direction could be from discrete options to a continuous selection, e.g., by means of a slider that the participants move over a continuous spectrum of answers. Here, however, it is questionable whether the participants of the experiment contain such fine granularity in their ethical judgment.

The ultimate goal of developing the experiment was to find a consensus in society that could guide the parameterization of the proposed principles. The results from Section 5.1 indicate that there is a wide variety of answers. To get a first indicator as a result of the survey, the ethical settings in Section 5.1 were obtained by taking the mean value over all participants. The mean value, however, would not constitute a consensus but rather a compromise since a consensus is defined as an opinion with which all members of a group agree. The search for a consensus in the future requires a reformulation of the questions. Hence, users should be

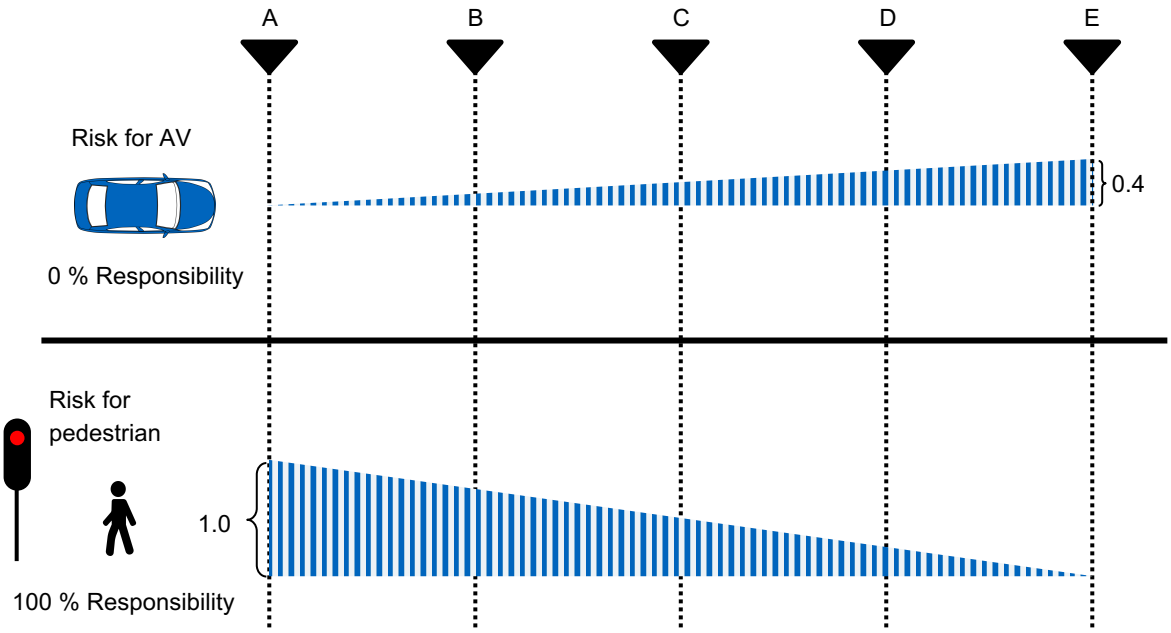


Figure 6.2: A promising future direction for the ethical experiment. In this example, the participants are asked to choose between five options (A-E) with two involved road users: the AV that is considered to be not responsible for a possible collision and a pedestrian that is held fully responsible, e.g., by crossing a red light. By shifting risks from the pedestrian to the AV, the total amount of risk can be reduced.

asked if they could agree to a given solution instead of asking for their view itself, not knowing if the opposite answer would also be acceptable to them. So the next task would be to find a setting that a majority - or, in the best case, everyone - can agree to. Therefore, the results of the study can serve as an initial point.

Analyzing each question separately gives valuable insight into its relevance. On the one hand, questions that receive diverse responses from participants and take significantly more time to answer can be considered important in revealing moral conflicts. On the other hand, questions where almost everyone agrees on one answer may not be as useful for future experiments. Figure 6.3 shows the distribution of answers for each question in the experiment. For example, question 5, which seeks to determine the weight given to the equality principle, generally receives low weightings for this principle from a majority of the participants. However, question 6, focusing on the trade-off between the Bayes and maximin principles, shows that participants have varying opinions, as there is no clear consensus on the answer. The complete list of the questions used in the experiment can be found in Appendix A.

6.3 Compliance with Legislation and Ethical Standards

Section 2.2.4 analyzed the current legal situation and ethical requirements, as well as guidelines. In this context, twelve requirements were extracted from three sources, namely the German ethics committee, the EU expert group, and the EU legislation of August 2022, for AV decision-making. In the following, these requirements will be discussed in light of the background of the presented algorithm and to what extent they are fulfilled.

R1: *The system must be designed in a way that dilemma situations do not appear.*

The German ethics committee defines a dilemma situation as a situation in which an AV is faced with the decision of having to necessarily realize one of two evils that must not be weighed. What must not be weighed

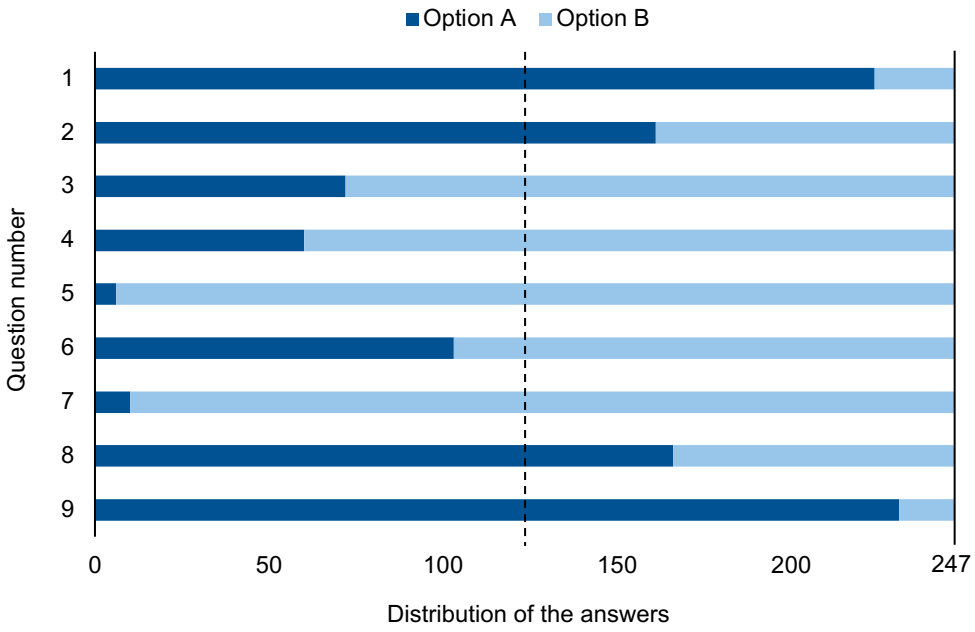


Figure 6.3: Distribution of answers to the questions of the ethical vehicle experiment. There was mostly agreement among respondents for some questions (e.g., Q5 or Q7), while other questions (e.g., Q6) had mixed answers.

will be discussed further within R3. The presented algorithm takes this requirement into account by not distinguishing between dilemma and non-dilemma situations. The consideration of a fair risk distribution takes place at any time. Situations in which, for example, the maximum accepted risk is exceeded could be interpreted as critical situations. In this case, safety is prioritized over comfort and mobility. The distribution of risk remains the same. Thus, according to the definition of the German ethics committee, these are not dilemma situations, as these situations are not characterized by a trade-off.

R2: *Higher priority of persons over animal or property damage.*

The harm model presented in Section 3.2.2 is purely based on personal damage. Property damage may only be considered implicitly by the correlation to personal damage. Thus, the desired prioritization is considered in the algorithm. Animals, however, are not considered in this work at all and could be part of future work.

R3: *Human lives must not be counted against each other, and bystanders must not be sacrificed.*

With reference to Article 1 of the German Basic Law, the ethics commission concludes that the sacrifice of innocent people in favor of other potential victims is inadmissible because the innocent would be degraded to a mere instrument and deprived of their subject quality [305]. At the same time, however, the principle of harm minimization is demanded. This can be reconciled with the principle of non-weighing since a probability prognosis is to be made from the situation in which the identity of the injured or killed (in contrast to the trolley cases) is not yet certain. In this respect, programming to minimize casualties (damage to property before personal injury, injury to persons before killing, least possible number of injured or killed) could at least be justified without violating Article 1 (1) of the Basic Law if the programming reduces the risk of each individual road user to the same extent. The risk assessment proposed in this thesis fulfills this requirement accordingly.

Secondly, the type of involvement (e.g., "bystander") in a situation is further emphasized here. Within the proposed work, the involvement is interpreted as the responsibility to risk emergence. As a consequence, road users who are not responsible for introducing risk into a situation can be assigned less risk as part of the risk distribution. Whether and to what extent this principle should be considered cannot be answered in

the context of this thesis since the principle of responsibility could not be taken into account in the empirical evaluation (Chapter 5).

R4: *Qualification according to personal characteristics (such as age and gender) is prohibited.*

The presented approach does not consider any personal characteristics, such as gender or age, in trajectory planning. However, implicit biases, such as a worse detection of underrepresented groups, are not taken into account here. They need to be considered in perception functionalities.

R5: *Redress inequalities in vulnerability among road users (#5).*

Inequalities in vulnerability are explicitly considered by the harm model in Section 3.2.2. The objective of harm estimation is to estimate the personal damage to every road user by considering their vulnerability. As a result, two parameter sets emerged for protected (vehicles, trucks, etc.) and unprotected (pedestrians, cyclists, etc.) road users. Furthermore, the masses are estimated and evaluated to account for differences in protection due to higher masses. In combination with the equality principle, which strives to distribute risk and, accordingly, potential harm equally among road users, this requirement is considered fulfilled.

R6: *Manage dilemmas by principles of risk distribution and shared ethical principles (#6).*

Chapter 3 investigated various principles for risk distribution and came to a similar conclusion as the EU expert group: a single ethical principle might not be sufficient to guide ethical decision-making for AVs. Although fictive, there are scenarios in every case where the usage of that specific single principle seems counterintuitive (Figure 3.9). Therefore, resolving trade-offs in AV decision-making by seeking a fair distribution of risk is a core component of this work. To constitute this risk distribution, various ethical principles, such as Bayes, equality, or maximin principle, are used in a shared manner here.

R7: *Enable user choice, seek informed consent options and develop related best practice industry standards (#8).*

This recommendation demands to "go beyond “take-it-or-leave-it” models of consent, to include agile and continuous consent options" [11]. The presented approach seeks consent options by means of a user survey (Chapter 4). However, this survey is designed as an apriori measurement and not as an agile instrument. A choice of different ethical settings could emerge in the future via the offerings of different manufacturers, as is already the case today. Yet, this work does not cover flexible, individualized, and continuous adaptation, which could be part of future research.

R8: *Promote a fair system for the attribution of moral and legal culpability for the behavior of CAVs (#18).*

This recommendation promotes a principle of responsibility to prevent both impunity for avoidable harm and scapegoating. Fair culpability attribution criteria are seen as key to reasonable moral and social practices of blame and punishment. Therefore, the ethical algorithm introduces the responsibility criteria. To what extent this way of assigning responsibility in (preventing) a collision is reasonable is still unclear. Also, in this work, this remains unanswered as these parameters are still kept open. However, the presented study offers a starting point to query proportionality in the eyes of society.

R9: *The ADS shall be able to detect the risk of collision with other road users.*

The detection and quantification of risks of potential collisions is a core component of this work. Equation 3.1 represents the two-dimensionality of risk consisting of collision probability and estimated harm. Considering further sources for collision risk next to prediction uncertainties should be the next step to improve the quality of risk assessment.

R10: *In the event of an unavoidable alternative risk to human life, the ADS shall not provide for any weighting on the basis of personal characteristics of humans.*

This requirement is similar to R4. Personal characteristics are not considered in the algorithm.

R11: *The protection of other human life outside the fully automated vehicle shall not be subordinated to the protection of human life inside the fully automated vehicle.*

This requirement represents one of the main reasons for the motivation of this work. As introduced in Chapter 1, the objective of this thesis was to find an alternative risk distribution based on ethical principles, which is in contrast to a selfish risk management strategy. As a result, the selfish principle was not considered one of the distribution principles. The remaining principles as part of the algorithm do not distinguish between the risks of the ego vehicle and the risks of third-party road users. The equality principle explicitly aims for the equality of all road users. However, there might be some implicit biases towards the ego vehicle that, e.g., emerge from the fact that the current state and plan of the ego-AV is well-known from the ego perspective in contrast to others. This information asymmetry was discussed in more detail in Section 6.2.3. The empiric evaluation underlined that as a consequence of the proposed risk distribution, the risk shifted from third-party road users, especially VRUs, to the ego-AV in comparison to a selfish algorithm. Thus, the effectiveness was proven in a trajectory planning algorithm.

R12: *The vulnerability of road users involved should be taken into account by the avoidance/mitigation strategy.*

This requirement from the EU legislation refers to R5 from the EU expert group and is met by the harm model as part of the algorithm. However, the general approach of this work was not to develop a strategy for emergency scenarios but to provide guidance based on ethical principles for all kinds of situations, which has been shown to be necessary. As this includes guidance for mitigation strategies, this requirement is considered satisfied.

6.4 Applicability in a Real-World Vehicle Application

An important requirement of this work was the applicability of the ethical algorithm for real-world use. Previous work has been limited to theoretical considerations and highly simplified simulations. The fact that a real algorithm was developed here, which forms part of the AV software, is, therefore, part of the novelty of this work. The extent to which this requirement has been met will be briefly discussed here. Since the real application in a vehicle was not shown within the scope of this work, the applicability, in theory, is to be evaluated on the basis of two essential aspects: Firstly, the ethical algorithm must meet the technical requirements of an AV trajectory planner. The second essential aspect is the computation time that results from the additional computational effort.

The algorithm, as presented, takes a CommonRoad map with dynamic obstacles and their probabilistic trajectory predictions as input and calculates a trajectory. The trajectory consists of locations in x and y with a rate of $10\ Hz$ and $2\ s$ planning horizon. By using polynomials of fourth (longitudinal) and fifth (lateral) order, only jerk-optimal trajectories are generated [272], which is supposed to ensure driveability in terms of their smoothness. Furthermore, the driveability is checked by kinematics checks using limits for acceleration or steering radius. The algorithmic interfaces are found in a similar way in related works that showed the applicability on a real vehicle, such as *Autoware* [306, 307]. Other approaches to the AV software distinguish between behavior or maneuver planning and trajectory planning as a subordinate task [158, 179]. A central element of the presented approach is the decision-making at the maneuver level, which cannot be separated from trajectory planning. Compatibility with such approaches would, therefore, only be given if the complete

planning task was replaced by the algorithm presented. In principle, the typical trajectory requirements of a controller are met by using jerk-optimal trajectories and kinematic checks. However, the next step would be to test the algorithm in interaction with a controller.

The algorithm was implemented as a prototype in the programming language Python. The software was not optimized explicitly for runtime and does not support multi-processing. Figure 6.4 provides a runtime analysis of the trajectory planning, as well as the required probabilistic prediction. It is noted that the runtime analysis was performed using a single core of an Intel Core i7 (9th generation) laptop Central Processing Unit (CPU) and no dedicated high-performance hardware. As the software is not parallelized, calculation time increases linearly with the number of sampled trajectories. The risk assessment accounts for $1.3\ ms$ per sampled trajectory, while the responsibility analysis adds another $0.3\ ms$ per trajectory. In the state-of-the-art baseline, the calculation time is around $1.1\ ms$ per trajectory. So, the functions regarding the ethical assessment of risk lead to an increase of $+145\ \%$ in computation time. However, the calculation of the ethical cost term only accounts for about $0.1\ \%$ of the additional computational effort. The majority consists of the calculation of risks, consisting of collision probability ($37.0\ \%$), harm model ($24.8\ \%$), and the evaluation with static obstacles, such as road boundaries ($23.7\ \%$). As the sampling of the trajectory has no dependencies between the sampling of different trajectories, there is great potential for parallelization to reduce runtime in the future.

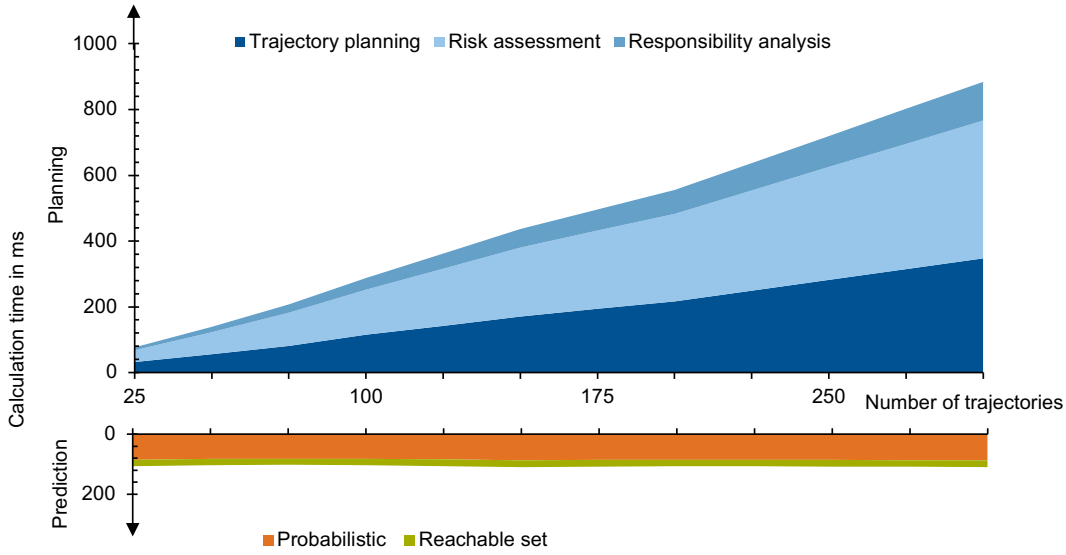


Figure 6.4: Runtime analysis of the proposed algorithm. Computation times are broken down separately into Prediction (bottom) and Planning (top) over the number of sampled trajectories. The proposed methods as extensions to the state of the art - namely the risk assessment and the responsibility analysis - together require about $2\ ms$ computing time per trajectory. The analysis was performed based on a prototype implementation without parallelization on a single thread of an Intel Core i7 (9th generation) laptop CPU. The figure was adapted from a previous author's publication [268]

The probabilistic prediction takes about $85\ ms$ of computation time, regardless of the number of trajectories sampled, and reachable set computation takes another $22\ ms$ in an average scenario. The representative scenario used here involves five road users that must be predicted with the corresponding method. For running the prediction without any other software in parallel, such as the trajectory planning algorithm, a computation time of $10\ ms$ is reported for a single road user on an NVIDIA V100 GPU [276]. The output of the uncertainties from the network as an additional element, important for ethical consideration, does not contribute to any measurable increase in computing time.

Although the implementation of the proposed algorithm is only in the form of a prototype, the method, in general, can be considered capable of running on a real AV. To this point, some further steps, such as the implementation in a real-time capable programming language like C++, are necessary. A potential conflict between ethical calculation and software performance remains an open point for further investigation. As a part of future work, the relevant functions for the ethical assessment will be implemented on the research vehicle at the Technical University of Munich [308], shown in Figure 6.5.

Figure 6.5: Research vehicle EDGAR [308], which is based on a Volkswagen T7 vehicle and has been extended by sensors and computing hardware to fulfill requirements for automated driving in urban environments with a safety driver.

6.5 Outlook

Ethics has been shown to play a crucial role in AV decision-making, and it is reasonable to consider it in algorithms. The core of this work presented an algorithm for ethical trajectory planning based on elaborated requirements. The algorithm should be seen as a first suggestion on how ethical aspects can be integrated into algorithms. Therefore, the discussion chapter has highlighted the remaining shortcomings and open points. However, it is important to acknowledge that AV providers can hardly be obligated to use a specific algorithm of this nature. Ultimately, it should become possible to incorporate ethics into decision-making processes without enforcing the utilization of a specific algorithm by OEMs.

Current legislation and recommendations on algorithmic decision-making lack concrete specifications and do not incorporate mechanisms to verify compliance with established guidelines. There are two potential approaches to address this issue: 1) implementing a reporting and disclosure obligation that outlines how ethical requirements are met and 2) employing a validation method, akin to test cases or scenarios, to demonstrate compliance with prescribed norms of behavior. These two approaches should be discussed below under the aspect of the presented work.

6.5.1 Reporting and Disclosure Obligation

Many AV providers, such as Waymo [309] or Cruise [310], publicly report on the safety concepts of their AVs, so far largely on a voluntary basis. A similar approach could be used to address the ethics of AVs. Requiring AV providers to report and disclose their software would allow regulators, researchers, and the

public to evaluate the algorithms and decision-making processes behind these vehicles. This transparency improves accountability and ensures that AV providers take responsibility for the performance and safety of their technology. By being transparent about their software and ethical considerations, OEMs can build trust among consumers, policymakers, and other stakeholders. A well-informed public is more likely to accept and embrace AVs, which can accelerate their adoption and potential benefits. However, requiring full disclosure of software could potentially expose valuable proprietary information to competitors, undermining the competitive advantage of AV providers. Protecting intellectual property is essential to fostering innovation and promoting continued advancements in autonomous driving technology.

Consequently, the question of the degree of abstraction in such reports is of high importance. Whether the underlying (ethical) values are declared during programming or even if the code is published open source makes a big difference. It should also be noted here that disclosure only works if the software is interpretable by humans. Approaches based on black-box algorithms, such as neural networks, would be unsuitable here.

In this context, the algorithm presented could serve as an inspiration for AV providers on how legal requirements could possibly be implemented.

6.5.2 Test Case Scenarios

A second option, in which no insight into the algorithm is required, is to systematically test the AV behavior in a set of scenarios. There are numerous examples of scenario-based testing from the point of view of safety argumentation [311]. For safety verification, scenarios are defined in which the AV software must meet certain metrics or quality criteria, such as no collisions. Similarly, this approach to safety can be extended to ethics. Therefore, instead of safety metrics, this requires the definition of metrics that reflect decision-making from an ethical point of view.

The algorithm presented here, which is motivated to act on the basis of ethical principles, can accordingly help identify such metrics and scenarios. A starting point could be the distribution principles presented. An approach to determining the ethical nature of a scenario is to establish a set of criteria. For instance, a scenario can be deemed ethical if a trade-off between discussed ethical principles - including the selfish principle - occurs. In that course, a library of ethical scenarios could be identified and used for proving a desired AV behavior. Further difficulties, next to the question of the right behavior, present the formulation of acceptance bounds. Since the legislation focuses on the algorithmic side and does not provide any information in terms of testing and verification, a formulation could be derived based on the algorithmic context. This requires the algorithmic implementation of the guidelines, as done in this work, to then define test cases based on these results.

7 Conclusion

Autonomous vehicles (AVs) hold the promise of revolutionizing transportation by enhancing safety, efficiency, and accessibility while reshaping our societal and urban landscapes. This thesis has delved into the multifaceted realm of AVs, recognizing that their impact extends far beyond technical advancement alone. Although technical aspects form a crucial component, it has been established that ethics plays an integral role in the successful integration of AVs onto public streets. As such, a holistic approach is imperative, calling for an interdisciplinary endeavor that adeptly balances ethical imperatives and technical demands.

Ethics and software development for AVs have often been treated as distinct domains in the existing literature. However, this work has illuminated the necessity for closer collaboration between these disciplines, driven by the inherent complexities of translating ethical principles into functional algorithms and navigating the regulatory landscape. The literature review has underscored the need for an algorithm that effectively incorporates ethics based on legislation, highlighting a significant gap in the current state of the art.

The trajectory planning method proposed in this thesis (Chapter 3) exemplifies a stride towards bridging this gap. The foundation of the algorithm in risk and uncertainty, aligned with risk ethics, paves the way for ethically guided decision-making. It has been shown that no single ethical principle accounts for the complexity of AV decision-making. Therefore, five ethical principles have been identified as relevant to guide the risk distribution in AV trajectory planning: The Bayes principle promotes trajectories that are expected to result in the overall lowest risk. The equality principle attains parity in risk distribution, while the maximin principle considers the moral claims of the worst-off. The responsibility principle adds the idea of a polluter-pays-principle in risk distribution, and the maximum acceptable risk pays attention to the emergence of the risk. However, parameterization of weights to these principles is identified as crucial for decision-making.

To face this challenge, a methodology is presented to ascertain these parameters, which follows the underlying idea of searching for a social consensus on how to distribute risks (Chapter 4). Based on the justice approach of Rawls, questions have been systematically designed, allowing for calculating weight values for each survey participant. At the culmination of the study, participants have the opportunity to compare their results with those of previous participants.

The so-called Ethical Vehicle Experiment, encompassing insights from 247 participants, underscores the multifaceted nature of ethical considerations. While a preference towards the Bayes principle emerged, the survey illuminated the diversity of moral perspectives, emphasizing the absence of a universal solution. The empirical evaluation of the ethical algorithm, guided by parameters derived from the survey, validated its effectiveness through simulations of 2,000 scenarios. The proposed algorithm, denoted as an ethical algorithm, was compared to an algorithm with selfish risk distribution and a state-of-the-art algorithm that does not consider risk. First, it could be shown that the explicit consideration of risk is beneficial for all road users since it significantly reduces the risk levels and corresponding accident harm. Second, using the ethical algorithm shifts risk from Vulnerable Road Users (VRUs) to the ego-AV. Therefore, the amount of risk reduction for VRUs is significantly higher than the risk increase for the controlled AV. This corresponds to the important legal demand of accounting for the vulnerability of VRUs.

Reviewing the proposed algorithm in the context of legislation showed that it basically fulfills all requirements and is thus seen as a major step forward. However, some open points, such as the complete parameterization as the result of a social consensus and not a compromise or considering more algorithmic uncertainties, remain. The algorithm is therefore seen as a first suggestion for implementing ethics in AVs, revealing some important insights from empirical analysis. Taking a closer look at the algorithm's interfaces and inference time showed that the proposed algorithm can principally run on a real AV, such as the research vehicle EDGAR. Therefore, integrating the proposed algorithm into a complete AV software running on real vehicles will be the next important step to move forward in bringing AVs to public streets that are ethically founded. The present research should, therefore, serve as a stimulus to include ethics into the AV software development, uncovering the potential of actively considering risks in the AV software.

List of Figures

List of Tables

Bibliography

[1] W. Gruel and J. M. Stanford, “Assessing the Long-term Effects of Autonomous Vehicles: A Speculative Approach,” *Transportation Research Procedia*, vol. 13, pp. 18–29, 2016, DOI: 10.1016/j.trpro.2016.05.003. Available: http://dx.doi.org/10.1016/j.trpro.2016.05.003.

[2] A. Pernestål and I. Kristoffersson, “Effects of driverless vehicles – Comparing simulations to get a broader picture,” *European Journal of Transport and Infrastructure Research*, vol. 19, no. 1, pp. 1–23, 2019, DOI: 10.18757/ejtir.2019.19.1.3944.

[3] National Transportation Safety Board Office of Highway Safety, “Vehicle Automation Report,” Tempe, AZ, 2019. Available: https://dms.ntsb.gov/pubdms/search/document.cfm?docID=477717&docketID=62978&mkey=96894.

[4] N. Goodall, “Ethical decision making during automated vehicle crashes,” *Transportation Research Record*, vol. 2424, no. 1, pp. 58–65, 2014, DOI: 10.3141/2424-07.

[5] P. Lin, “Why Ethics Matters for Autonomous Cars,” in *Autonomous Driving: Technical, Legal and Social Aspects* 2016, pp. 69–85, ISBN: 9783662488478. DOI: 10.1007/978-3-662-48847-8.

[6] A. Kriebitz, R. Max and C. Lütge, “The German Act on Autonomous Driving : Why Ethics Still Matters,” *Philosophy & Technology*, pp. 1–13, 2022, DOI: 10.1007/s13347-022-00526-2. Available: https://doi.org/10.1007/s13347-022-00526-2.

[7] A. K. Faulhaber, A. Dittmer, F. Blind, M. A. Wächter, S. Timm, et al., “Human Decisions in Moral Dilemmas are Largely Described by Utilitarianism: Virtual Car Driving Study Provides Guidelines for Autonomous Driving Vehicles,” *Science and Engineering Ethics*, vol. 25, no. 2, pp. 399–418, 2019, DOI: 10.1007/s11948-018-0020-x. Available: https://doi.org/10.1007/s11948-018-0020-x.

[8] P. Foot, “The Problem of Abortion and the Doctrine of the Double Effect,” *Oxford Review*, vol. 5, 1967.

[9] R. Hu. “*Ethical Murder: The Trolley Dilemma*,” 2023. Available: https://theaxiom.ca/ethical-murder-the-trolley-dilemma/.

[10] S. Karnouskos, “Self-Driving Car Acceptance and the Role of Ethics,” *IEEE Transactions on Engineering Management*, vol. 67, no. 2, pp. 252–265, 2020, DOI: 10.1109/TEM.2018.2877307.

[11] Horizon 2020 Commission Expert Group, *Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*, Publication Office of the European Union: Luxembourg, 2020, ISBN: 978-92-76-17867-5. DOI: 10.2777/035239.

[12] Ethik-Komission, “Automatisiertes und Vernetztes Fahren (Bundesministerium für Verkehr und digitale Infrastruktur),” Bundesministerium für Verkehr und digitale Infrastruktur, 2017. Available: https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile.

[13] European Commission, “COMMISSION IMPLEMENTING REGULATION (EU) 2022/1426,” 2022.

[14] J.-F. Bonnefon, A. Shariff and I. Rahwan, “The Trolley, The Bull Bar, and Why Engineers Should Care About The Ethics of Autonomous Cars,” *Proceedings of the IEEE*, vol. 107, no. 3, pp. 502–504, 2019, DOI: 10.1109/JPROC.2019.2897447. Available: https://ieeexplore.ieee.org/document/8662742/.

[15] J. F. Bonnefon, A. Shariff and I. Rahwan, “The social dilemma of autonomous vehicles,” *Science*, vol. 352, no. 6293, pp. 1573–1576, 2016, DOI: 10.1126/science.aaf2654.

[16] N. J. Goodall, “From trolleys to risk: Models for ethical autonomous driving,” *American Journal of Public Health*, vol. 107, no. 4, p. 496, 2017, DOI: 10.2105/AJPH.2017.303672.

[17] SAE International, “SAE J3016: Levels of Driving Automation,” *Society of Automotive Engineers*, p. 202104, 2021. Available: https://www.sae.org/standards/content/j3016_202104/.

[18] C. Katrakazas, M. Quddus, W. H. Chen and L. Deka, “Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions,” *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 416–442, 2015, DOI: 10.1016/j.trc.2015.09.011. Available: http://dx.doi.org/10.1016/j.trc.2015.09.011.

[19] D. S. Carlson, K. M. Kacmar and L. L. Wadsworth, “The Impact of Moral Intensity Dimensions on Ethical Decision-Making: Assessing the Relevance of Orientation,” *Journal of Managerial*, vol. 21, no. 4, pp. 534–551, 2009. Available: http://www.jstor.org/stable/40604668.

[20] J. Rawls, *A Theory of Justice*, Harvard University Press, Cambridge, 1971.

[21] S. O. Hansson, M. Å. Belin and B. Lundgren, “Self-Driving Vehicles—an Ethical Overview,” *Philosophy and Technology*, vol. 34, no. 4, pp. 1383–1408, 2021, DOI: 10.1007/s13347-021-00464-5. Available: https://doi.org/10.1007/s13347-021-00464-5.

[22] R. Jenkins, “Autonomous Vehicles - Ethics & Law: Toward an Overlapping Consensus,” no. September, p. 32, 2016. Available: https://na-production.s3.amazonaws.com/documents/AV-Ethics-Law.pdf.

[23] M. Alawadhi, J. Almazrouie, M. Kamil and K. A. Khalil, “Review and analysis of the importance of autonomous vehicles liability: a systematic literature review,” *International Journal of System Assurance Engineering and Management*, vol. 11, no. 6, pp. 1227–1249, 2020, DOI: 10.1007/s13198-020-00978-9. Available: https://doi.org/10.1007/s13198-020-00978-9.

[24] D. Sickert, “Ethics in Autonomous Driving,” *ResearchGate*, 2019. Available: https://www.researchgate.net/publication/332370194.

[25] Y. Pala, “Autonomous Vehicle Technology Effect on the Unskilled Labor Economy,” 2023, DOI: 10.13140/RG.2.2.10724.04486.

[26] N. Reed, T. Leiman, P. Palade, M. Martens and L. Kester, “Ethics of automated vehicles: breaking traffic rules for road safety,” *Ethics and Information Technology*, vol. 23, no. 4, pp. 777–789, 2021, DOI: 10.1007/s10676-021-09614-x. Available: https://doi.org/10.1007/s10676-021-09614-x.

[27] K. Kim, J. S. Kim, S. Jeong, J.-H. Park and H. K. Kim, “Cybersecurity for autonomous vehicles: Review of attacks and defense,” *Computers & Security*, vol. 103, p. 102150, 2021, DOI: 10.1016/j.cose.2020.102150. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167404820304235.

[28] R. Dave, E. R Sowells Boone and K. Roy, “Efficient Data Privacy and Security in Autonomous Cars,” *Journal of Computer Sciences and Applications*, vol. 7, no. 1, pp. 31–36, 2019, DOI: 10.12691/jcsa-7-1-5.

[29] S. Atakishiyev, M. Salameh, H. Yao and R. Goebel, “Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions,” 2021. Available: http://arxiv.org/abs/2112.11561.

[30] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman and A. Galstyan, “A Survey on Bias and Fairness in Machine Learning,” *ACM Computing Surveys*, vol. 54, no. 6, 2021, DOI: 10.1145/3457607.

[31] B. Wilson, J. Hoffman and J. Morgenstern, “Predictive Inequity in Object Detection,” *ArXiv*, 2019. Available: http://arxiv.org/abs/1902.11097.

[32] I. Zliobaite, “A survey on measuring indirect discrimination in machine learning,” *arxiv.org/abs/1511.00148*, vol. 0, no. 0, 2015. Available: http://arxiv.org/abs/1511.00148.

[33] Audi AG, “&AudiSocAlty-Study: Autonomous Driving on the Road to Social Acceptance,” pp. 1–74, 2021. Available: https://www.audi.com/content/dam/gbp2/company/research/audi-beyond/2021/AUDI_SocAITy_Study_dgtl_1201_English_small.pdf.

[34] F. Henze, D. Fasbender and C. Stiller, “How Can Automated Vehicles Explain Their Driving Decisions? Generating Clarifying Summaries Automatically,” *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2022-June, pp. 935–942, 2022, ISBN: 9781665488211. DOI: 10.1109/IV51971.2022.9827197.

[35] N. J. Goodall, “Machine Ethics and Automated Vehicles,” in *Road Vehicle Automation* 2014, pp. 93–102, ISBN: 978-3-319-05989-1. DOI: 10.1007/978-3-319-05990-7_9. Available: https://link.springer.com/chapter/10.1007/978-3-319-05990-7_9.

[36] T. Fraichard, “Will the Driver Seat Ever Be Empty ?,” RR-8493, INRIA, 2014.

[37] R. Benenson, T. Fraichard and M. Parent, “Achievable safety of driverless ground vehicles,” in *2008 10th International Conference on Control, Automation, Robotics and Vision, ICARCV 2008*, 2008, pp. 515–521, ISBN: 9781424422876. DOI: 10.1109/ICARCV.2008.4795572.

[38] T. Fraichard and J. J. Kuffner. “*Guaranteeing motion safety for robots*,” Apr. 2012. DOI: 10.1007/s10514-012-9278-z. Available: https://link.springer.com/article/10.1007/s10514-012-9278-z.

[39] H. Wang, Y. Huang, A. Khajepour, Y. Zhang, Y. Rasekhipour, et al., “Crash Mitigation in Motion Planning for Autonomous Vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3313–3323, 2019, ISBN: 9781668436950. DOI: 10.1109/TITS.2018.2873921. Available: https://ieeexplore.ieee.org/document/8617711/.

[40] A. Bautin, L. Martinez-Gomez and T. Fraichard, “Inevitable collision states: A probabilistic perspective,” in *Proceedings - IEEE International Conference on Robotics and Automation*, 2010, pp. 4022–4027, ISBN: 9781424450381. DOI: 10.1109/ROBOT.2010.5509233.

[41] International Organization for Standardization, “ISO 26262: Road vehicles - Functional safety,” 2018.

[42] J. C. Gerdes, “The Virtues of Automated Vehicle Safety - Mapping Vehicle Safety Approaches to Their Underlying Ethical Frameworks,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2020, pp. 107–113, ISBN: 9781728166735. DOI: 10.1109/IV47402.2020.9304583.

[43] T. Fournier, “Will My Next Car Be a Libertarian or a Utilitarian: Who Will Decide?,” *IEEE Technology and Society Magazine*, vol. 35, no. 2, pp. 40–45, 2016, DOI: 10.1109/MTS.2016.2554441.

[44] S. Applin, “Autonomous Vehicle Ethics: Stock or custom?,” *IEEE Consumer Electronics Magazine*, vol. 6, no. 3, pp. 108–110, 2017, DOI: 10.1109/MCE.2017.2684917.

[45] D. Martin, “Who Should Decide How Machines Make Morally Laden Decisions?,” *Science and Engineering Ethics*, vol. 23, no. 4, pp. 951–967, 2017, DOI: 10.1007/s11948-016-9833-7.

[46] M. Brandão, “Moral autonomy and equality of opportunity for algorithms in autonomous vehicles,” *Frontiers in Artificial Intelligence and Applications*, vol. 311, pp. 302–310, 2018, ISBN: 9781614999300. DOI: 10.3233/978-1-61499-931-7-302.

[47] G. Contissa, F. Lagioia and G. Sartor, “The Ethical Knob: ethically-customisable automated vehicles and the law,” *Artificial Intelligence and Law*, vol. 25, no. 3, pp. 365–378, 2017, DOI: 10.1007/s10506-017-9211-z.

[48] J. Gogoll and J. F. Müller, “Autonomous Cars: In Favor of a Mandatory Ethics Setting,” *Science and Engineering Ethics*, vol. 23, no. 3, pp. 681–700, 2017, DOI: 10.1007/s11948-016-9806-x.

[49] J. De Freitas, A. Censi, B. W. Smith, L. D. Lillo, S. E. Anthony, et al., “From driverless dilemmas to more practical commonsense tests for automated vehicles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 118, no. 11, pp. 1–9, 2021, DOI: 10.1073/pnas.2010202118.

[50] R. Davnall, “Solving the Single-Vehicle Self-Driving Car Trolley Problem Using Risk Theory and Vehicle Dynamics,” *Science and Engineering Ethics*, 2019, ISBN: 1194801900. DOI: 10.1007/s11948-019-00102-6. Available: https://doi.org/10.1007/s11948-019-00102-6.

[51] M. Geisslinger, F. Poszler, J. Betz, C. Lütge and M. Lienkamp, “Autonomous Driving Ethics: from Trolley Problem to Ethics of Risk,” *Philosophy & Technology*, 2021, DOI: 10.1007/s13347-021-00449-4. Available: https://doi.org/10.1007/s13347-021-00449-4%20https://link.springer.com/10.1007/s13347-021-00449-4.

[52] A. Wischnewski, M. Euler, S. Gümüs and B. Lohmann, “Tube model predictive control for an autonomous race car,” *Vehicle System Dynamics*, 2021, DOI: 10.1080/00423114.2021.1943461. Available: https://doi.org/10.1080/00423114.2021.1943461.

[53] A. Etzioni and O. Etzioni, “Incorporating Ethics into Artificial Intelligence,” *Journal of Ethics*, vol. 21, no. 4, pp. 403–418, 2017, DOI: 10.1007/s10892-017-9252-2.

[54] L. Zaho and W. Li, “"Choose for No Choose" - Random Selecting Option for the Trolley Problem in Autonomous Driving,” *9th IEEE International Conference on Logistics, Informatics and Service Sciences*, 2021, ISBN: 9789811556814.

[55] T. Gill, “Ethical dilemmas are really important to potential adopters of autonomous vehicles,” *Ethics and Information Technology*, vol. 23, no. 4, pp. 657–673, 2021, DOI: 10.1007/s10676-021-09605-y. Available: https://doi.org/10.1007/s10676-021-09605-y.

[56] A. Shariff, J. F. Bonnefon and I. Rahwan, “Psychological roadblocks to the adoption of self-driving vehicles,” *Nature Human Behaviour*, vol. 1, no. 10, pp. 694–696, 2017, DOI: 10.1038/s41562-017-0202-6.

[57] J. J. Thomson, “The Trolley Problem,” *The Yale Law Journal*, vol. 94, no. 6, pp. 1395–1415, 1985.

[58] S. Nyholm, “The ethics of crashes with self-driving cars: A roadmap, I,” *Philosophy Compass*, vol. 13, no. 7, e12507, 2018, DOI: 10.1111/phc3.12507.

[59] A. Wolkenstein, “What has the Trolley Dilemma ever done for us (and what will it do in the future)? On some recent debates about the ethics of self-driving cars,” *Ethics and Information Technology*, vol. 20, no. 3, pp. 163–173, 2018, DOI: 10.1007/s10676-018-9456-6. Available: https://doi.org/10.1007/s10676-018-9456-6.

[60] L. R. Sütfeld, P. König and G. Pipa, “Towards a Framework for Ethical Decision Making in Automated Vehicles,” *PsyArXiv*, pp. 1–27, 2019. Available: https://psyarxiv.com/4duca/.

[61] S. Motwani, T. Sharma and A. Gupta, “Ethics in Autonomous Vehicle Software: The Dilemmas,” *Computer*, vol. 54, no. 8, pp. 46–55, 2021, DOI: 10.1109/MC.2021.3077576.

[62] I. Coca-Vila, “Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law,” *Criminal Law and Philosophy*, vol. 12, no. 1, pp. 59–82, 2018, DOI: 10.1007/s11572-017-9411-3.

[63] G. Keeling, “Why Trolley Problems Matter for the Ethics of Automated Vehicles,” *Science and Engineering Ethics*, 2019, ISBN: 1194801900. DOI: 10.1007/s11948-019-00096-1. Available: https://doi.org/10.1007/s11948-019-00096-1.

[64] S. Nyholm and J. Smids, “The Ethics of Accident-Algorithms for Self-Driving Cars: an Applied Trolley Problem?,” *Ethical Theory and Moral Practice*, vol. 19, no. 5, pp. 1275–1289, 2016, DOI: 10.1007/s10677-016-9745-2.

[65] V. Bhargava and T. W. Kim, “Autonomous vehicles and moral uncertainty,” *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, no. January, pp. 5–19, 2017, ISBN: 9780190652951. DOI: 10.1093/oso/9780190652951.003.0001.

[66] A. Kauppinen, “Who Should Bear the Risk When Self-Driving Vehicles Crash?,” *Journal of Applied Philosophy*, vol. 38, no. 4, pp. 630–645, 2021, DOI: 10.1111/japp.12490. Available: https://onlinelibrary.wiley.com/doi/10.1111/japp.12490.

[67] H. Y. Liu, “Irresponsibilities, inequalities and injustice for autonomous vehicles,” *Ethics and Information Technology*, vol. 19, no. 3, pp. 193–207, 2017, DOI: 10.1007/s10676-017-9436-2.

[68] J. Himmelreich, “Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations,” *Ethical Theory and Moral Practice*, vol. 21, no. 3, pp. 669–684, 2018, DOI: 10.1007/s10677-018-9896-4.

[69] H. Zhao, K. Dimovitz, B. Staveland and L. Medsker, “Responding to challenges in the design of moral autonomous vehicles,” *AAAI Fall Symposium - Technical Report*, vol. FS-16-01 -, pp. 169–173, 2016, ISBN: 9781577357759.

[70] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, et al., “The Moral Machine experiment,” *Nature*, vol. 563, no. 7729, pp. 59–64, 2018, DOI: 10.1038/s41586-018-0637-6. Available: http://dx.doi.org/10.1038/s41586-018-0637-6.

[71] A. K. Frison, P. Wintersberger and A. Riener, “First person trolley problem: Evaluation of drivers’ ethical decisions in a driving simulator,” *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Adjunct Proceedings*, pp. 117–122, 2016, ISBN: 9781450346542. DOI: 10.1145/3004323.3004336.

[72] A. G. Mirnig and A. Meschtscherjakov, “Trolled by the trolley problem on what maters for ethical decision making in automated vehicles,” *Conference on Human Factors in Computing Systems - Proceedings*, no. Chi, pp. 1–10, 2019, ISBN: 9781450359702. DOI: 10.1145/3290605.3300739.

[73] H. M. Roff, “The folly of trolleys: Ethical challenges and autonomous vehicles,” *Brookings*, no. Kagan 1989, 2018. Available: https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/.

[74] N. J. Goodall, “Away from Trolley Problems and Toward Risk Management,” *Applied Artificial Intelligence*, vol. 30, no. 8, pp. 810–821, 2016, DOI: 10.1080/08839514.2016.1229922. Available: http://dx.doi.org/10.1080/08839514.2016.1229922.

[75] L. Alexander and M. Moore, “Deontological Ethics,” in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, ed. Metaphysics Research Lab, Stanford University, 2021.

[76] I. Asimov, *I, Robot*, Greenwich, CT, Fawcett Publications, 1950.

[77] T. M. Powers, “Prospects for a kantian machine,” *IEEE Intelligent Systems*, vol. 21, no. 4, pp. 46–51, 2006, DOI: 10.1109/MIS.2006.77.

[78] S. Shalev-Shwartz, S. Shammah and A. Shashua, “On a Formal Model of Safe and Scalable Self-driving Cars,” *arxiv.org/abs/1708.06374*, pp. 1–37, 2017. Available: http://arxiv.org/abs/1708.06374.

[79] C. Shea-Blymyer and H. Abbas, “Algorithmic ethics: Formalization and verification of autonomous vehicle obligations,” *ACM Transactions on Cyber-Physical Systems*, vol. 5, no. 4, 2021, DOI: 10.1145/3460975.

[80] H. Wang, Y. Huang, A. Khajepour, D. Cao and C. Lv, “Ethical Decision-Making Platform in Autonomous Vehicles with Lexicographic Optimization Based Model Predictive Controller,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8164–8175, 2020, DOI: 10.1109/TVT.2020.2996954.

[81] S. M. Thornton, S. Pan, S. M. Erlien and J. C. Gerdes, “Incorporating Ethical Considerations into Automated Vehicle Control,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1429–1439, 2017, DOI: 10.1109/TITS.2016.2609339.

[82] S. Pan, S. M. Thornton and J. C. Gerdes, “Prescriptive and proscriptive moral regulation for autonomous vehicles in approach and avoidance,” *2016 IEEE International Symposium on Ethics in Engineering, Science and Technology, ETHICS 2016*, 2016, ISBN: 9781509023172. DOI: 10.1109/ETHICS.2016.7560049.

[83] A. Censi, K. Slutsky, T. Wongpiromsarn, D. Yershov, S. Pendleton, et al., “Liability, ethics, and culture-aware behavior specification using rulebooks,” *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2019-May, pp. 8536–8542, 2019, ISBN: 9781538660263. DOI: 10.1109/ICRA.2019.8794364.

[84] F. Lindner, R. Mattmüller and B. Nebel, “Evaluation of the moral permissibility of action plans,” *Artificial Intelligence*, vol. 287, p. 103350, 2020, DOI: 10.1016/j.artint.2020.103350. Available: https://doi.org/10.1016/j.artint.2020.103350.

[85] L. Millan-Blanquel, S. M. Veres and R. C. Purshouse, “Ethical considerations for a decision making system for autonomous vehicles during an inevitable collision,” *2020 28th Mediterranean Conference on Control and Automation, MED 2020*, pp. 514–519, 2020, ISBN: 9781728157429. DOI: 10.1109/MED48518.2020.9183263.

[86] L. Dennis, M. Fisher, M. Slavkovik and M. Webster, “Formal verification of ethical choices in autonomous systems,” *Robotics and Autonomous Systems*, vol. 77, pp. 1–14, 2016, DOI: 10.1016/j.robot.2015.11.012.

[87] J. Loh, “Roboterethik. Über eine noch junge Bereichsethik,” *Information Philosophie*, pp. 20–33, 2017.

[88] N. J. Goodall, “Can you program ethics into a self-driving car?,” *IEEE Spectrum*, vol. 53, no. 6, 2016, DOI: 10.1109/MSPEC.2016.7473149.

[89] C. Brändle and M. W. Schmidt, “Autonomous Driving and Public Reason: a Rawlsian Approach,” *Philosophy and Technology*, vol. 34, no. 4, pp. 1475–1499, 2021, DOI: 10.1007/s13347-021-00468-1. Available: https://doi.org/10.1007/s13347-021-00468-1.

[90] M. Dietrich and T. H. Weisswange, “Distributive justice as an ethical principle for autonomous vehicle behavior beyond hazard scenarios,” *Ethics and Information Technology*, vol. 21, no. 3, pp. 227–239, 2019, DOI: 10.1007/s10676-019-09504-3. Available: https://doi.org/10.1007/s10676-019-09504-3%20http://link.springer.com/10.1007/s10676-019-09504-3.

[91] D. Leben, “A Rawlsian algorithm for autonomous vehicles,” *Ethics and Information Technology*, vol. 19, no. 2, pp. 107–115, 2017, DOI: 10.1007/s10676-017-9419-3. Available: http://link.springer.com/10.1007/s10676-017-9419-3.

[92] G. Keeling, *Against Leben’s Rawlsian Collision Algorithm for Autonomous Vehicles*. vol. 44, Springer International Publishing, pp. 259–272, 2018, ISBN: 9783319964485. DOI: 10.1007/978-3-319-96448-5_29. Available: http://dx.doi.org/10.1007/978-3-319-96448-5_29.

[93] W. Sinnott-Armstrong, “Consequentialism,” in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta and U. Nodelman, ed. Metaphysics Research Lab, Stanford University, 2022.

[94] C. Bartneck, C. Lütge, A. Wagner and S. Welsh, *Ethik in KI und Robotik*, Carl Hanser Verlag GmbH Co KG., 2019.

[95] A. Johnsen, N. Strand, J. Andersson, C. Patten, C. Kraetsch, et al., “Literature review on the acceptance and road safety, ethical, legal, social and economic implications of automated vehicles,” Institut für empirische Soziologie an der Universität Erlangen-Nürnberg., 2018. Available: https://www.researchgate.net/publication/325786957.

[96] J. E. Pickering, M. Podsiadly and K. J. Burnham, “A Model-to-Decision Approach for the Autonomous Vehicle (AV) Ethical Dilemma: AV Collision with a Barrier/Pedestrian(s),” *IFAC-PapersOnLine*, vol. 52, no. 8, pp. 381–386, 2019, DOI: 10.1016/j.ifacol.2019.08.080. Available: https://doi.org/10.1016/j.ifacol.2019.08.080.

[97] W. Kumfer and R. Burgess, “Investigation into the role of rational ethics in crashes of automated vehicles,” *Transportation Research Record*, vol. 2489, pp. 130–136, 2015, DOI: 10.3141/2489-15.

[98] C. Luetge, “The German Ethics Code for Automated and Connected Driving,” *Philosophy and Technology*, vol. 30, no. 4, pp. 547–558, 2017, DOI: 10.1007/s13347-017-0284-0.

[99] D. Hübner and L. White, “Crash Algorithms for Autonomous Cars: How the Trolley Problem Can Move Us Beyond Harm Minimisation,” *Ethical Theory and Moral Practice*, vol. 21, no. 3, pp. 685–698, 2018, DOI: 10.1007/s10677-018-9910-x. Available: https://doi.org/10.1007/s10677-018-9910-x.

[100] J. C. Gerdes and S. M. Thornton, “Implementable Ethics for Autonomous Vehicles,” in *Autonomes Fahren*. vol. 9403 Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 87–102, ISBN: 978-3-662-45853-2. DOI: 10.1007/978-3-662-45854-9_5. Available: http://link.springer.com/10.1007/978-3-662-45854-9_5.

[101] V. Schäffner, “Caught up in ethical dilemmas: An adapted consequentialist perspective on self-driving vehicles,” *Frontiers in Artificial Intelligence and Applications*, vol. 311, pp. 327–335, 2018, ISBN: 9781614999300. DOI: 10.3233/978-1-61499-931-7-327.

[102] N. Berberich and K. Diepold, “The Virtuous Machine - Old Ethics for New Technology?,” *arXiv*, pp. 1–25, 2018. Available: http://arxiv.org/abs/1806.10322.

[103] S. M. Thornton, “Autonomous Vehicle Motion Planning With Ethical Considerations,” PhD thesis, Stanford University, 2018.

[104] M. Bansal, A. Krizhevsky and A. Ogale, “ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst,” in *Robotics: Science and Systems 2019*, 2018, pp. 1–20. Available: http://arxiv.org/abs/1812.03079.

[105] X. Wang, H. Krasowski and M. Althoff, “CommonRoad-RL: A Configurable Reinforcement Learning Environment for Motion Planning of Autonomous Vehicles,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2021-Septe, pp. 466–472, 2021, ISBN: 9781728191423. DOI: 10.1109/ITSC48978.2021.9564898.

[106] P. Kulicki, M. P. Musielewicz and R. Trypuz, “Virtue Ethics for Autonomous Cars (short version),” *ResearchGate Preprint*, no. May, 2019, DOI: 10.13140/RG.2.2.32982.50241.

[107] B. Rath, *Entscheidungstheorien der Risikoethik*, Tectum Wissenschaftsverlag, 2011, ISBN: 978-3-8288-2682-3.

[108] N. Goodall, “More than trolleys plausible, ethically ambiguous scenarios likely to be encountered by automated vehicles,” *Transfers*, vol. 9, no. 2, pp. 45–58, 2019, DOI: 10.3167/TRANS.2019.090204.

[109] S. O. Hansson, “Risk,” in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta and U. Nodelman, ed. Metaphysics Research Lab, Stanford University, 2022.

[110] G. Keeling, K. Evans, S. M. Thornton, G. Mecacci and F. Santoni de Sio, “Four Perspectives on What Matters for the Ethics of Automated Vehicles,” in *Road Vehicle Automation 6* 2019, pp. 49–60, ISBN: 9783030229337. DOI: 10.1007/978-3-030-22933-7_6. Available: http://link.springer.com/10.1007/978-3-030-22933-7_6.

[111] G. Persad, A. Wertheimer and E. J. Emanuel, “Principles for allocation of scarce medical interventions,” *The Lancet*, vol. 373, no. 9661, pp. 423–431, 2009, DOI: 10.1016/S0140-6736(09)60137-9. Available: http://dx.doi.org/10.1016/S0140-6736(09)60137-9.

[112] S. O. Hansson, “Ethics and radiation protection,” *Journal of Radiological Protection*, vol. 27, no. 2, pp. 147–156, 2007, DOI: 10.1088/0952-4746/27/2/002.

[113] M. A. Islam and S. I. Rashid, “Algorithm for ethical decision making at times of accidents for autonomous vehicles,” *4th International Conference on Electrical Engineering and Information and Communication Technology, iCEEiCT 2018*, pp. 438–442, 2019, ISBN: 9781538682791. DOI: 10.1109/CEEICT.2018.8628155.

[114] R. Johansson and J. Nilsson, “Disarming the Trolley Problem – Why Self-driving Cars do not Need to Choose Whom to Kill,” *4th International Workshop on Critical Automotive Applications: Robustness & Safety (CARS 2016)*, 2016.

[115] K. Evans, N. de Moura, S. Chauvier, R. Chatila and E. Dogan, “Ethical Decision Making in Autonomous Vehicles: The AV Ethics Project,” *Science and Engineering Ethics*, vol. 26, no. 6, pp. 3285–3312, 2020, DOI: 10.1007/s11948-020-00272-8. Available: https://doi.org/10.1007/s11948-020-00272-8.

[116] J. A. Cervantes, S. López, L. F. Rodríguez, S. Cervantes, F. Cervantes, et al., *Artificial Moral Agents: A Survey of the Current Status*. vol. 26, Springer Netherlands, pp. 501–532, 2020, ISBN: 1194801900. DOI: 10.1007/s11948-019-00151-x. Available: https://doi.org/10.1007/s11948-019-00151-x.

[117] J. Rhim, J.-H. Lee, M. Chen, A. Lim, M. Magnusson, et al., “A Deeper Look at Autonomous Vehicle Ethics: An Integrative Ethical Decision-Making Framework to Explain Moral Pluralism,” vol. 8, 2021, DOI: 10.3389/frobt.2021.632394.

[118] H. Etienne, “A practical role-based approach for autonomous vehicle moral dilemmas,” *Big Data and Society*, vol. 9, no. 2, 2022, DOI: 10.1177/20539517221123305.

[119] D.-A. Frank, P. Chrysochou, P. Mitkidis and D. Ariely, “Human decision-making biases in the moral dilemmas of autonomous vehicles,” *Scientific Reports*, vol. 9, no. 1, pp. 1–19, 2019, DOI: 10.1038/s41598-019-49411-7.

[120] M. M. Mayer, R. Bell and A. Buchner, “Self-protective and self-sacrificing preferences of pedestrians and passengers in moral dilemmas involving autonomous vehicles,” *PLoS ONE*, vol. 16, no. 12 December, 2021, DOI: 10.1371/journal.pone.0261673. Available: https://doi.org/10.1371/journal.pone.0261673.

[121] W. Hugemann, “Driver Reaction Times in Road Traffic,” *11th EVU Conference*, vol. 29, pp. 1–12, 2002.

[122] L. R. Sütfeld, R. Gast, P. König and G. Pipa, “Using virtual reality to assess ethical decisions in road traffic scenarios: Applicability of value-of-life-based models and influences of time pressure,” *Frontiers in Behavioral Neuroscience*, vol. 11, no. October 2015, pp. 1–13, 2017, DOI: 10.3389/fnbeh.2017.00122.

[123] S. Samuel, S. Yahoodik, Y. Yamani, K. Valluru and D. L. Fisher, “Ethical decision making behind the wheel – A driving simulator study,” *Transportation Research Interdisciplinary Perspectives*, vol. 5, p. 100147, 2020, DOI: 10.1016/j.trip.2020.100147. Available: https://doi.org/10.1016/j.trip.2020.100147.

[124] P. Liu and J. Liu, “Selfish or Utilitarian Automated Vehicles? Deontological Evaluation and Public Acceptance,” *International Journal of Human-Computer Interaction*, vol. 37, no. 13, pp. 1231–1242, 2021, DOI: 10.1080/10447318.2021.1876357. Available: https://doi.org/10.1080/10447318.2021.1876357.

[125] P. Wintersberger, A. K. Prison, A. Riener and S. Hasirlioglu, “The experience of ethics: Evaluation of self harm risks in automated vehicles,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2017, pp. 385–391, ISBN: 9781509048045. DOI: 10.1109/IVS.2017.7995749.

[126] L. T. Bergmann, L. Schlicht, C. Meixner, P. König, G. Pipa, et al., “Autonomous vehicles require socio-political acceptance—an empirical and philosophical perspective on the problem of moral decision making,” *Frontiers in Behavioral Neuroscience*, vol. 12, p. 31, 2018, DOI: 10.3389/fnbeh.2018.00031. Available: www.frontiersin.org.

[127] C. M. de Melo, S. Marsella and J. Gratch, “Risk of Injury in Moral Dilemmas With Autonomous Vehicles,” *Frontiers in Robotics and AI*, vol. 7, no. January, pp. 1–10, 2021, DOI: 10.3389/frobt.2020.572529.

[128] B. Meder, N. Fleischhut, N. C. Krumnau and M. R. Waldmann, “How Should Autonomous Cars Drive? A Preference for Defaults in Moral Judgments Under Risk and Uncertainty,” *Risk Analysis*, vol. 39, no. 2, pp. 295–314, 2019, DOI: 10.1111/risa.13178.

[129] J. Robinson, J. Smyth, R. Woodman and V. Donzella, “Ethical considerations and moral implications of autonomous vehicles and unavoidable collisions,” *Theoretical Issues in Ergonomics Science*, vol. 23, no. 4, pp. 435–452, 2022, DOI: 10.1080/1463922X.2021.1978013. Available: https://doi.org/10.1080/1463922X.2021.1978013%20https://www.tandfonline.com/doi/full/10.1080/1463922X.2021.1978013.

[130] G. M. Grasso, C. Lucifora, P. Perconti and A. Plebe, *Integrating human acceptable morality in autonomous vehicles*. vol. 1131 AISC, Springer International Publishing, pp. 41–45, 2020, ISBN: 9783030395117. DOI: 10.1007/978-3-030-39512-4_7. Available: http://dx.doi.org/10.1007/978-3-030-39512-4_7.

[131] C. Lucifora, G. M. Grasso, P. Perconti and A. Plebe, “Moral dilemmas in self-driving cars,” *Rivista Internazionale di Filosofia e Psicologia*, vol. 11, no. 2, pp. 238–250, 2020, DOI: 10.4453/rifp.2020.0015.

[132] C. Lucifora, ·. Giorgio, M. Grasso, P. Perconti and A. Plebe, “Moral reasoning and automatic risk reaction during driving,” *Cognition, Technology & Work*, vol. 23, pp. 705–713, 2021, DOI: 10.1007/s10111-021-00675-y. Available: https://doi.org/10.1007/s10111-021-00675-y.

[133] M. Black, “The Gap Between "Is" and "Should",” *The Philosophical Review*, vol. 73, no. 2, p. 165, 1964, DOI: 10.2307/2183334. Available: https://www.jstor.org/stable/2183334?origin=crossref.

[134] S. Krügel and M. Uhl, “Autonomous vehicles and moral judgments under risk,” *Transportation Research Part A: Policy and Practice*, vol. 155, pp. 1–10, 2022, DOI: 10.1016/j.tra.2021.10.016. Available: https://linkinghub.elsevier.com/retrieve/pii/S0965856421002718.

[135] J. Savulescu, G. Kahane and C. Gyngell, “From public preferences to ethical policy,” *Nature Human Behaviour*, vol. 3, no. 12, pp. 1241–1243, 2019, DOI: 10.1038/s41562-019-0711-6. Available: http://dx.doi.org/10.1038/s41562-019-0711-6.

[136] O. Shigeharu, O. Yumi, K. Keigo and M. Mizumoto, “Egocentric, Altruistic, or Hypocritic?: A Cross-Cultural Study of Choice between Pedestrian-first and Driver-first of Autonomous Car,” *IEEE Access*, no. 2021, 2023, DOI: https://doi.org/10.36227/techrxiv.16688941.v3.

[137] UNESCO, “Recommendation on the Ethics of Artificial Intelligence,” rep. November, 2022. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[138] High-Level Expert Group on Artificial Intelligence (AI HLEG), “Ethics Guidelines for Trustworthy AI,” 11/2019. Available: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.

[139] European Commission, “Artificial Intelligence Act COM(2021) 206 final,” *European Commission*, vol. 0106, pp. 1–108, 2021.

[140] Z. Ramadan, “The gamification of trust: the case of China’s “social credit”,” *Marketing Intelligence and Planning*, vol. 36, no. 1, pp. 93–107, 2018, DOI: 10.1108/MIP-06-2017-0100.

[141] K. Bothmer and T. Schlippe, “Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-Based Recommendation System,” in *Lecture Notes in Networks and Systems*, 2023, pp. 69–80, ISBN: 9783031215681. DOI: 10.1007/978-3-031-21569-8_7. Available: https://link.springer.com/chapter/10.1007/978-3-031-21569-8_7.

[142] A. Jobin, M. Ienca and E. Vayena, “The global landscape of AI ethics guidelines,” *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019, ISBN: 4225601900. DOI: 10.1038/s42256-019-0088-2.

[143] Google, “Recommendations for Regulating AI,” Alphabet Inc., 2020. Available: https://ai.google/static/documents/recommendations-for-regulating-ai.pdf.

[144] H. Domin, J. Van Dodick, C. Lawrence and F. Rossi, “Standards for protecting at-risk groups in AI bias auditing,” IBM, rep. November, 2022.

[145] Bundesministerium für Digitales und Verkehr, “Verordnung zur Regelung des Betriebs von Kraftfahrzeugen mit automatisierter und autonomer Fahrfunktion und zur Änderung straßenverkehrsrechtlicher Vorschriften,” 2022.

[146] D. Coelho and M. Oliveira, “A Review of End-to-End Autonomous Driving in Urban Environments,” *IEEE Access*, vol. 10, no. June, pp. 75296–75311, 2022, DOI: 10.1109/ACCESS.2022.3192019.

[147] A. Tampuu, T. Matiisen, M. Semikin, D. Fishman and N. Muhammad, “A Survey of End-to-End Driving: Architectures and Training Methods,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 4, pp. 1364–1384, 2022, DOI: 10.1109/TNNLS.2020.3043505.

[148] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, et al., “End-to-end Autonomous Driving: Challenges and Frontiers,” no. JUNE, 2023. Available: http://arxiv.org/abs/2306.16927.

[149] P. S. Chib and P. Singh, “Recent Advancements in End-to-End Autonomous Driving using Deep Learning: A Survey,” no. June, pp. 1–21, 2023. Available: http://arxiv.org/abs/2307.04370.

[150] A. Srivastava, “Sense-Plan-Act in Robotic Applications,” no. February 2019, 2019, DOI: 10.13140/RG.2.2.21308.36481. Available: https://www.researchgate.net/publication/349248621.

[151] S. D. Pendleton, H. Andersen, X. Du, X. Shen, M. Meghjani, et al., “Perception, planning, control, and coordination for autonomous vehicles,” *Machines*, vol. 5, no. 1, pp. 1–54, 2017, DOI: 10.3390/machines5010006.

[152] J. Betz, T. Betz, F. Fent, M. Geisslinger, A. Heilmeier, et al., “TUM Autonomous Motorsport: An Autonomous Racing Software for the Indy Autonomous Challenge,” *Journal of Field Robotics*, 2023, DOI: 10.1002/ROB.22153. Available: https://onlinelibrary.wiley.com/doi/10.1002/rob.22153%20http://arxiv.org/abs/2205.15979.

[153] P. Karle and M. Lienkamp. “*Lecture 01 : Introduction to Autonomous Driving*,” Munich, Germany, 2021.

[154] J. Wei, J. M. Snider, T. Gu, J. M. Dolan and B. Litkouhi, “A behavioral planning framework for autonomous driving,” *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 458–464, 2014, ISBN: 9781479936380. DOI: 10.1109/IVS.2014.6856582.

[155] K. Esterle, T. Kessler and A. Knoll, “Optimal Behavior Planning for Autonomous Driving: A Generic Mixed-Integer Formulation,” *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 1914–1921, 2020, ISBN: 9781728166735. DOI: 10.1109/IV47402.2020.9304743.

[156] T. Guy, J. M. Dolan and J. W. Lee, “Automated tactical maneuver discovery, reasoning and trajectory planning for autonomous driving,” *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Novem, pp. 5474–5480, 2016, ISBN: 9781509037629. DOI: 10.1109/IROS.2016.7759805.

[157] B. Brumitt, A. Stentz and M. Hebert, “Autonomous Driving with Concurrent Goals and Multiple Vehicles: Mission Planning and Architecture,” *Autonomous Robots*, vol. 12, no. 2, pp. 135–156, 2002, DOI: 10.1023/A:1014008325793.

[158] L. Claussmann, M. Revilloud, D. Gruyer and S. Glaser, “A Review of Motion Planning for Highway Autonomous Driving,” *IEEE Transactions on Intelligent Transportation Systems*, no. May, pp. 1–23, 2019, DOI: 10.1109/tits.2019.2913998.

[159] S. Manzinger, M. Koschi and M. Althoff, “CommonRoad: Cost Function Specification,” *Technische Universität München*, pp. 2–5, 2020.

[160] A. Rizaldi and M. Althoff, “Formalising Traffic Rules for Accountability of Autonomous Vehicles,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2015-Octob, pp. 1658–1665, 2015, ISBN: 9781467365956. DOI: 10.1109/ITSC.2015.269.

[161] J. Ziegler, P. Bender, T. Dang and C. Stiller, “Trajectory planning for Bertha - A local, continuous method,” *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 450–457, 2014, ISBN: 9781479936380. DOI: 10.1109/IVS.2014.6856581.

[162] S. Magdici and M. Althoff, “Fail-safe motion planning of autonomous vehicles,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 452–458, 2016, ISBN: 9781509018895. DOI: 10.1109/ITSC.2016.7795594.

[163] D. Kim, H. Peng, S. Bai and J. M. Maguire, “Control of integrated powertrain with electronic throttle and automatic transmission,” *IEEE Transactions on Control Systems Technology*, vol. 15, no. 3, pp. 474–482, 2007, DOI: 10.1109/TCST.2007.894641.

[164] W. Xu, J. Wei, J. M. Dolan, H. Zhao and H. Zha, “A real-time motion planner with trajectory optimization for autonomous vehicles,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2061–2067, 2012, ISBN: 9781467314039. DOI: 10.1109/ICRA.2012.6225063.

[165] J. C. Hayward, “Near-miss determination through use of scale of danger,” *Highway Research Board*, no. 384, pp. 24–35, 1972. Available: http://onlinepubs.trb.org/Onlinepubs/hrr/1972/384/384-004.pdf.

[166] S. M. Mahmud, L. Ferreira, M. S. Hoque and A. Tavassoli, “Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs,” *IATSS Research*, vol. 41, no. 4, pp. 153–163, 2017, DOI: 10.1016/j.iatssr.2017.02.001. Available: https://doi.org/10.1016/j.iatssr.2017.02.001.

[167] J. Hillenbrand, A. M. Spieker and K. Kroschel, “A multilevel collision mitigation approach - Its situation assessment, decision making, and performance tradeoffs,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 4, pp. 528–540, 2006, DOI: 10.1109/TITS.2006.883115.

[168] S. Wagner, K. Groh, T. Kuhbeck, M. Dörfel and A. Knoll, “Using Time-to-React based on Naturalistic Traffic Object Behavior for Scenario-Based Risk Assessment of Automated Driving,” *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2018-June, no. Iv, pp. 1521–1528, 2018, ISBN: 9781538644522. DOI: 10.1109/IVS.2018.8500624.

[169] Y. Kuang, X. Qu and S. Wang, “A tree-structured crash surrogate measure for freeways,” *Accident Analysis and Prevention*, vol. 77, pp. 137–148, 2015, DOI: 10.1016/j.aap.2015.02.007. Available: http://dx.doi.org/10.1016/j.aap.2015.02.007.

[170] V. Astarita, G. Guido, A. Vitale and V. Giofré, “A new microsimulation model for the evaluation of traffic safety performances,” *European Transport - Trasporti Europei*, no. 51, pp. 1–16, 2012.

[171] A. Eidehall and L. Petersson, “Statistical threat assessment for general road scenes using Monte Carlo sampling,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 137–147, 2008, DOI: 10.1109/TITS.2007.909241.

[172] M. Althoff, O. Stursberg and M. Buss, “Model-based probabilistic collision detection in autonomous driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 299–310, 2009, DOI: 10.1109/TITS.2009.2018966.

[173] Y. Lin and M. Althoff, “CommonRoad-CriMe: A Toolbox for Criticality Measures of Autonomous Vehicles,” in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–8, DOI: 10.1109/IV55152.2023.10186673. Available: https://ieeexplore.ieee.org/document/10186673/.

[174] R. Trauth, M. Kaufeld, M. Geisslinger and J. Betz, “Learning and Adapting Behavior of Autonomous Vehicles through Inverse Reinforcement Learning,” in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–8, DOI: 10.1109/IV55152.2023.10186668.

[175] S. Rosbach, V. James, S. Grosjohann, S. Homoceanu and S. Roth, “Driving with Style: Inverse Reinforcement Learning in General-Purpose Planning for Automated Driving,” *IEEE International Conference on Intelligent Robots and Systems*, pp. 2658–2665, 2019, ISBN: 9781728140049. DOI: 10.1109/IROS40897.2019.8968205.

[176] S. Teng, X. Hu, P. Deng, B. Li, Y. Li, et al., “Motion Planning for Autonomous Driving: The State of the Art and Future Perspectives,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 6, pp. 3692–3711, 2023, DOI: 10.1109/TIV.2023.3274536.

[177] H. Caesar, J. Kabzan, K. S. Tan, W. K. Fong, E. Wolff, et al., “NuPlan: A closed-loop ML-based planning benchmark for autonomous vehicles,” 2021. Available: http://arxiv.org/abs/2106.11810.

[178] W. Schwarting, J. Alonso-Mora and D. Rus, “Planning and Decision-Making for Autonomous Vehicles,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 187–210, 2018, DOI: 10.1146/annurev-control-060117-105157.

[179] B. Paden, M. Cap, S. Z. Yong, D. Yershov and E. Frazzoli, “A Survey of Motion Planning and Control Techniques for Self-driving Urban Vehicles,” pp. 1–27, 2016. Available: http://arxiv.org/abs/1604.07446.

[180] D. González, J. Pérez, V. Milanés and F. Nashashibi, “A Review of Motion Planning Techniques for Automated Vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1135–1145, 2016, DOI: 10.1109/TITS.2015.2498841.

[181] T. Stahl, A. Wischnewski, J. Betz and M. Lienkamp, “Multilayer Graph-Based Trajectory Planning for Race Vehicles in Dynamic Scenarios,” *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2–7, 2019.

[182] D. Ferguson, T. M. Howard and M. Likhachev, “Motion planning in urban environments: Part II,” *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pp. 1070–1076, 2008, ISBN: 9781424420582. DOI: 10.1109/IROS.2008.4651124.

[183] D. Ferguson, T. M. Howard and M. Likhachev, “Motion planning in urban environments,” *Springer Tracts in Advanced Robotics*, vol. 56, pp. 61–89, 2009, ISBN: 9783642039904. DOI: 10.1007/978-3-642-03991-1_2.

[184] M. Pivtoraiko, R. A. Knepper and A. Kelly, “Differentially constrained mobile robot motion planning in state lattices,” *Journal of Field Robotics*, vol. 26, no. 3, pp. 308–333, 2009, DOI: 10.1002/rob.20285. Available: http://onlinelibrary.wiley.com/doi/10.1002/rob.21514/abstract%20http://doi.wiley.com/10.1002/rob.20285.

[185] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959, DOI: 10.1007/BF01386390. Available: http://link.springer.com/10.1007/BF01386390.

[186] A. Stentz, “The D * Algorithm for Real-Time Planning of Optimal Traverses,” no. September, p. 30, 1994.

[187] A. Stentz, “The Focussed D* Algorithm for Real-Time Replanning,” no. August, 1995.

[188] M. Rowold, L. Ögretmen, T. Kerbl and B. Lohmann, “Efficient Spatiotemporal Graph Search for Local Trajectory Planning on Oval Race Tracks,” *Actuators*, vol. 11, no. 11, pp. 1–18, 2022, DOI: 10.3390/act11110319.

[189] M. Rowold, L. Ögretmen, U. Kasolowsky and B. Lohmann, “Online Time-Optimal Trajectory Planning on Three-Dimensional Race Tracks,” 2023. Available: http://arxiv.org/abs/2304.10954.

[190] S. M. LaValle and J. J. Kuffner, “Randomized Kinodynamic Planning,” *The International Journal of Robotics Research*, vol. 20, no. 5, pp. 378–400, 2001, DOI: 10.1177/02783640122067453. Available: http://journals.sagepub.com/doi/10.1177/02783640122067453.

[191] S. Karaman and E. Frazzoli, “Sampling-based algorithms for optimal motion planning,” *International Journal of Robotics Research*, vol. 30, no. 7, pp. 846–894, 2011, DOI: 10.1177/0278364911406761.

[192] E. Frazzoli, M. A. Dahleh and E. Feron, “Real-time motion planning for agile autonomous vehicles,” *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 1, pp. 116–129, 2002, ISBN: 9781563479786. DOI: 10.2514/2.4856.

[193] L. Ma, J. Xue, K. Kawabata, J. Zhu, C. Ma, et al., “Efficient sampling-based motion planning for on-road autonomous driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1961–1976, 2015, DOI: 10.1109/TITS.2015.2389215.

[194] L. I. Reyes Castro, P. Chaudhari, J. Tümová, S. Karaman, E. Frazzoli, et al., “Incremental sampling-based algorithm for minimum-violation motion planning,” *Proceedings of the IEEE Conference on Decision and Control*, pp. 3217–3224, 2013, ISBN: 9781467357173. DOI: 10.1109/CDC.2013.6760374.

[195] L. E. Kavraki, P. Švestka, J. C. Latombe and M. H. Overmars, “Probabilistic roadmaps for path planning in high-dimensional configuration spaces,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 4, pp. 566–580, 1996, DOI: 10.1109/70.508439.

[196] D. Hsu and Z. Sun, “Adaptively combining multiple sampling strategies for probabilistic roadmap planning,” *2004 IEEE Conference on Robotics, Automation and Mechatronics*, vol. 2, pp. 774–779, 2004, ISBN: 0780386469. DOI: 10.1109/ramech.2004.1438016.

[197] S. M. LaValle, M. S. Branicky and S. R. Lindemann, “On the relationship between classical grid search and probabilistic roadmaps,” *International Journal of Robotics Research*, vol. 23, no. 7-8, pp. 673–692, 2004, DOI: 10.1177/0278364904045481.

[198] X. Li, Z. Sun, Q. Zhu and D. Liu, “A unified approach to local trajectory planning and control for autonomous driving along a reference path,” *2014 IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2014*, pp. 1716–1721, 2014, ISBN: 9781479939787. DOI: 10.1109/ICMA.2014.6885959.

[199] T. Hesse, D. Hess and T. Sattel, “Motion planning for passenger vehicles - Force field trajectory optimization for automated driving,” *Proceedings of the IASTED International Conference on Robotics and Applications*, no. June 2016, pp. 284–292, 2010, ISBN: 9780889868625. DOI: 10.2316/P.2010.706-066.

[200] M. Werling, S. Kammel, J. Ziegler and L. Gröll, “Optimal trajectories for time-critical street scenarios using discretized terminal manifolds,” *International Journal of Robotics Research*, vol. 31, no. 3, pp. 346–359, 2012, DOI: 10.1177/0278364911423042.

[201] M. Werling, *Ein neues Konzept für die Trajektoriengenerierung und -stabilisierung in zeitkritischen Verkehrsszenarien*. vol. 60, pp. 53–54, 2012, ISBN: 9783866446311. DOI: 10.1524/auto.2012.0970.

[202] P. Falcone, F. Borrelli, J. Asgari, H. E. Tseng and D. Hrovat, “Predictive active steering control for autonomous vehicle systems,” *IEEE Transactions on Control Systems Technology*, vol. 15, no. 3, pp. 566–580, 2007, DOI: 10.1109/TCST.2007.894653.

[203] A. Liniger, A. Domahidi and M. Morari, “Optimization-based autonomous racing of 1:43 scale RC cars,” *Optimal Control Applications and Methods*, vol. 36, no. 5, pp. 628–647, 2015, DOI: 10.1002/oca.2123.

[204] W. Schwarting, J. Alonso-Mora, L. Pauli, S. Karaman and D. Rus, “Parallel autonomy in automated vehicles: Safe motion generation with minimal intervention,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 1928–1935, 2017, ISBN: 9781509046331. DOI: 10.1109/ICRA.2017.7989224.

[205] J. K. Subosits and J. C. Gerdes, “From the Racetrack to the Road: Real-Time Trajectory Replanning for Autonomous Driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 309–320, 2019, DOI: 10.1109/tiv.2019.2904390.

[206] D. Dolgov, S. Thrun, M. Montemerlo and J. Diebel, “Path planning for autonomous vehicles in unknown semi-structured environments,” *International Journal of Robotics Research*, vol. 29, no. 5, pp. 485–501, 2010, DOI: 10.1177/0278364909359210.

[207] A. Wischnewski, M. Geisslinger, J. Betz, T. Betz, F. Fent, et al., “Indy Autonomous Challenge - Autonomous Race Cars at the Handling Limits,” in *12th International Munich Chassis Symposium 2021* Springer Berlin Heidelberg, 2022, pp. 1–16.

[208] M. Naumann and C. Stiller, *AIB-MDP: Continuous Probabilistic Motion Planning for Automated Vehicles by Leveraging Action Independent Belief Spaces*, IEEE, pp. 6373–6380, 2022, ISBN: 9781665479271. DOI: 10.1109/IROS47612.2022.9981696. Available: https://ieeexplore.ieee.org/document/9981696/.

[209] F. Henze, D. Fabbender and C. Stiller, “Sensitivity Analysis of a Planning Algorithm Considering Uncertainties,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2020, pp. 1128–1134, DOI: 10.1109/IV47402.2020.9304575.

[210] F. Henze, D. Fabender and C. Stiller, “Identifying Admissible Uncertainty Bounds for the Input of Planning Algorithms,” *IEEE Transactions on Intelligent Vehicles*, 2021, DOI: 10.1109/TIV.2021.3119352.

[211] D. Althoff, J. J. Kuffner, D. Wollherr and M. Buss, “Safety assessment of robot trajectories for navigation in uncertain and dynamic environments,” *Autonomous Robots*, vol. 32, no. 3, pp. 285–302, 2012, DOI: 10.1007/s10514-011-9257-9.

[212] A. Eidehall and L. Petersson, “Threat assessment for general road scenes using Monte Carlo sampling,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 9, no. 1, pp. 1173–1178, 2006, ISBN: 1424400945. DOI: 10.1109/itsc.2006.1707381.

[213] A. Lambert, D. Grayer, G. S. Pierre and A. N. Ndieng, “Collision probability assessment for speed control,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pp. 1043–1048, 2008, DOI: 10.1109/ITSC.2008.4732692.

[214] O. S. Tas and C. Stiller, “Tackling Existence Probabilities of Objects with Motion Planning for Automated Urban Driving,” *Robotics: Science and Systems*, 2020, DOI: 10.48550/arxiv.2002.01254. Available: https://arxiv.org/abs/2002.01254v2%20http://arxiv.org/abs/2002.01254.

[215] Ö. Ş. Taş and C. Stiller, “Limited Visibility and Uncertainty Aware Motion Planning for Automated Driving,” *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2018-June, no. Iv, pp. 1171–1178, 2018, ISBN: 9781538644522. DOI: 10.1109/IVS.2018.8500369.

[216] L. Wang, C. Burger and C. Stiller, “Reasoning about Potential Hidden Traffic Participants by Tracking Occluded Areas,” in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2021, pp. 157–163, ISBN: 9781728191423. DOI: 10.1109/ITSC48978.2021.9564584.

[217] C. Hubmann, M. Becker, D. Althoff, D. Lenz and C. Stiller, “Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles,” *IEEE Intelligent Vehicles Symposium, Proceedings*, 2017, ISBN: 9781509048045. DOI: 10.1109/IVS.2017.7995949.

[218] J. I. Ge, B. Schürmann, R. M. Murray and M. Althoff, “Risk-aware motion planning for automated vehicle among human-driven cars,” pp. 3987–3993, 2019, ISBN: 9781538679289. Available: https://www.cds.caltech.edu/~murray/preprints/gsma19-acc_s.pdf.

[219] R. T. Rockafellar and S. Uryasev, “Optimization of Conditional Value-at-Risk,” *Journal of Risk*, 2000.

[220] L. Wang, C. F. Lopez and C. Stiller, “Realistic Single-Shot and Long-Term Collision Risk for a Human-Style Safer Driving,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2020, pp. 2073–2080, DOI: 10.1109/IV47402.2020.9304541.

[221] P. Chaudhari, T. Wongpiromsarny and E. Frazzoli, “Incremental minimum-violation control synthesis for robots interacting with external agents,” *Proceedings of the American Control Conference*, pp. 1761–1768, 2014, ISBN: 9781479932726. DOI: 10.1109/ACC.2014.6859284.

[222] D. Kamran, C. F. Lopez, M. Lauer and C. Stiller, “Risk-Aware High-level Decisions for Automated Driving at Occluded Intersections with Reinforcement Learning,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2020, pp. 1205–1212, DOI: 10.1109/IV47402.2020.9304606.

[223] W. Xu, J. Pan, J. Wei and J. M. Dolan, “Motion planning under uncertainty for on-road autonomous driving,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2507–2512, 2014, DOI: 10.1109/ICRA.2014.6907209.

[224] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960, DOI: 10.1115/1.3662552.

[225] Z. Huang, H. Liu, J. Wu, W. Huang and C. Lv, “Learning Interaction-aware Motion Prediction Model for Decision-making in Autonomous Driving,” 2023. Available: http://arxiv.org/abs/2302.03939.

[226] J. Mänttäri, “Interpretable , Interaction-Aware Vehicle Trajectory Prediction with Uncertainty Interpretable , Interaction-Aware Vehicle Trajectory Prediction with Uncertainty,” PhD thesis, 2021, ISBN: 978-91-7873-770-3.

[227] S. Lefèvre, D. Vasquez and C. Laugier, “A survey on motion prediction and risk assessment for intelligent vehicles,” *ROBOMECH Journal*, vol. 1, no. 1, pp. 1–14, 2014, DOI: 10.1186/s40648-014-0001-z.

[228] P. Karle, M. Geisslinger, J. Betz and M. Lienkamp, “Scenario Understanding and Motion Prediction for Autonomous Vehicles - Review and Comparison,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 16962–16982, 2022, DOI: 10.1109/TITS.2022.3156011. Available: https://www.ieee.org/publications/rights/index.html.

[229] M. Tsogas, A. Polychronopoulos and A. Amditis, “Unscented Kalman filter design for curvilinear motion models suitable for automotive safety applications,” *2005 7th International Conference on Information Fusion, FUSION*, vol. 2, pp. 1295–1302, 2005, ISBN: 0780392868. DOI: 10.1109/ICIF.2005.1592006.

[230] R. Miller and Q. Huang, “An adaptive peer-to-peer collision warning system,” *IEEE Vehicular Technology Conference*, vol. 1, pp. 317–321, 2002, ISBN: 0780374843. DOI: 10.1109/VTC.2002.1002718.

[231] S. Ammoun and F. Nashashibi, “Real time trajectory prediction for collision risk estimation between vehicles,” *Proceedings - 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing, ICCP 2009*, pp. 417–422, 2009, ISBN: 9781424450077. DOI: 10.1109/ICCP.2009.5284727.

[232] N. Kaempchen, K. Weiss, M. Schaefer and K. C. Dietmayer, “IMM object tracking for high dynamic driving maneuvers,” *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 825–830, 2004, ISBN: 0780383109. DOI: 10.1109/ivs.2004.1336491.

[233] M. Althoff and J. M. Dolan, “Online verification of automated road vehicles using reachability analysis,” *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, 2014, DOI: 10.1109/TRO.2014.2312453.

[234] M. Koschi, C. Pek, M. Beikirch and M. Althoff, “Set-Based Prediction of Pedestrians in Urban Environments Considering Formalized Traffic Rules,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2018-Novem, pp. 2704–2711, 2018, ISBN: 9781728103235. DOI: 10.1109/ITSC.2018.8569434.

[235] S. Maierhofer, P. Moosbrugger and M. Althoff, “Formalization of Intersection Traffic Rules in Temporal Logic,” *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2022-June, no. Iv, pp. 1135–1144, 2022, ISBN: 9781665488211. DOI: 10.1109/IV51971.2022.9827153.

[236] M. Naumann, F. Wirth, F. Oboril, K. U. Scholl, M. S. Elli, et al., “On responsibility sensitive safety in car-following situations - A parameter analysis on german highways,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2021, pp. 83–90, ISBN: 9781728153940. DOI: 10.1109/IV48863.2021.9575420.

[237] J. Yoshida, “Robotaxi Priorities : Avoid Crashes or Avoid Blame ?,” 2022. Available: https://ojoyoshidareport.com/robotaxi-priorities-avoid-crashes-or-avoid-blame/?utm_source=rss&utm_medium=rss&utm_campaign=robotaxi-priorities-avoid-crashes-or-avoid-blame.

[238] G. R. De Campos, R. Kianfar and M. Brannstrom, “Precautionary Safety for Autonomous Driving Systems: Adapting Driving Policies to Satisfy Quantitative Risk Norms,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2021-Septe, pp. 645–652, 2021, ISBN: 9781728191423. DOI: 10.1109/ITSC48978.2021.9564879.

[239] M. Althoff, O. Stursberg and M. Buss, “Stochastic reachable sets of interacting traffic participants,” *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 1086–1092, 2008, ISBN: 9781424425693. DOI: 10.1109/IVS.2008.4621131.

[240] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings and A. Mouzakitis, “Deep Learning-based Vehicle Behaviour Prediction For Autonomous Driving Applications: A Review,” pp. 1–13, 2019. Available: http://arxiv.org/abs/1912.11676.

[241] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, et al., “nuScenes: A multimodal dataset for autonomous driving,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, no. June, pp. 11621–11631, 2020.

[242] M. F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, et al., “Argoverse: 3D tracking and forecasting with rich maps,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 8740–8749, 2019, ISBN: 9781728132938. DOI: 10.1109/CVPR.2019.00895.

[243] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, et al., “Human Motion Trajectory Prediction: A Survey,” *Journal of Vibration and Control*, p. 107754631982824, 2019. Available: http://arxiv.org/abs/1905.06113.

[244] T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom and E. M. Wolff, “CoverNet: Multimodal Behavior Prediction using Trajectory Sets,” 2019. Available: http://arxiv.org/abs/1911.10298.

[245] A. Zyner, S. Worrall, J. Ward and E. Nebot, “Long short term memory for driver intent prediction,” *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 1484–1489, 2017, ISBN: 9781509048045. DOI: 10.1109/IVS.2017.7995919.

[246] F. Altché, A. D. L. Fortelle, F. Altché, A. D. La, F. An, et al., “An LSTM Network for Highway Trajectory Prediction To cite this version : HAL Id : hal-01691832 An LSTM Network for Highway Trajectory Prediction,” 2018.

[247] N. Deo and M. M. Trivedi, “Convolutional social pooling for vehicle trajectory prediction,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2018-June, pp. 1549–1557, 2018, ISBN: 9781538661000. DOI: 10.1109/CVPRW.2018.00196.

[248] R. Chandra, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, et al., “Forecasting Trajectory and Behavior of Road-Agents Using Spectral Clustering in Graph-LSTMs,” 2019. Available: http://arxiv.org/abs/1912.01118.

[249] A. Hammam, S. E. Ghobadi, F. Bonarens and C. Stiller, “Real-time Uncertainty Estimation Based on Intermediate Layer Variational Inference,” in *Proceedings - CSCS 2021: ACM Computer Science in Cars Symposium*, 2021, ISBN: 9781450391399. DOI: 10.1145/3488904.3493381. Available: https://dl.acm.org/doi/10.1145/3488904.3493381.

[250] A. Hammam, F. Bonarens, S. E. Ghobadi and C. Stiller, “Predictive Uncertainty Quantification of Deep Neural Networks using Dirichlet Distributions,” in *Proceedings - CSCS 2022: 6th ACM Computer Science in Cars Symposium*, 2022, ISBN: 9781450397865. DOI: 10.1145/3568160.3570233. Available: https://dl.acm.org/doi/10.1145/3568160.3570233.

[251] J. Mercat, T. Gilles, N. E. Zoghby, G. Sandou, D. Beauvois, et al., “Multi-head attention for multi-modal joint vehicle motion forecasting,” *arXiv*, pp. 9638–9644, 2019, ISBN: 9781728173955.

[252] C. Choi, A. Patil and S. Malla, “DROGON: A Causal Reasoning Framework for Future Trajectory Forecast,” 2019. Available: http://arxiv.org/abs/1908.00024.

[253] C. Choi, “Shared Cross-Modal Trajectory Prediction for Autonomous Driving,” no. i, 2020. Available: http://arxiv.org/abs/2004.00202.

[254] N. Rhinehart, K. M. Kitani and P. Vernaza, “R2P2: A reparameterized pushforward policy for diverse, precise generative path forecasting,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11217 LNCS, pp. 794–811, 2018, ISBN: 9783030012601. DOI: 10.1007/978-3-030-01261-8_47.

[255] H. Cui, V. Radosavljevic, F. C. Chou, T. H. Lin, T. Nguyen, et al., “Multimodal trajectory predictions for autonomous driving using deep convolutional networks,” *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2019-May, pp. 2090–2096, 2019, ISBN: 9781538660263. DOI: 10.1109/ICRA.2019.8793868.

[256] N. Djuric, V. Radosavljevic, H. Cui, T. Nguyen, F.-C. Chou, et al., “Uncertainty-aware Short-term Motion Prediction of Traffic Actors for Autonomous Driving,” 2018. Available: http://arxiv.org/abs/1808.05819.

[257] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, et al., “Planning-based prediction for pedestrians,” *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, vol. 1, pp. 3931–3936, 2009, ISBN: 9781424438044. DOI: 10.1109/IROS.2009.5354147.

[258] S. Rosbach, V. James, S. Grosjohann, S. Homoceanu, X. Li, et al., “Driving Style Encoder: Situational Reward Adaptation for General-Purpose Planning in Automated Driving,” *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 6419–6425, 2020, ISBN: 9781728173955. DOI: 10.1109/ICRA40945.2020.9196778.

[259] Y. Wu, J. Hou, G. Chen and A. Knoll, “Trajectory Prediction Based on Planning Method Considering Collision Risk,” *ICARM 2020 - 2020 5th IEEE International Conference on Advanced Robotics and Mechatronics*, pp. 466–470, 2020, ISBN: 9781728164793. DOI: 10.1109/ICARM49381.2020.9195282.

[260] D. Li and J. Du, “Maximum Entropy Inverse Reinforcement Learning Based on Behavior Cloning of Expert Examples,” *Proceedings of 2021 IEEE 10th Data Driven Control and Learning Systems Conference, DDCLS 2021*, pp. 996–1000, 2021, ISBN: 9781665424233. DOI: 10.1109/DDCLS52934.2021.9455476.

[261] J. Ho and S. Ermon, “Generative Adversarial Imitation Learning,” in *New Journal of Physics*, 2016.

[262] J. Fu, K. Luo and S. Levine, “Learning robust rewards with adversarial inverse reinforcement learning,” *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pp. 1–15, 2018.

[263] J. Basl and J. Behrends, “Why Everyone Has It Wrong About the Ethics of Autonomous Vehicles,” in *Frontiers of Engineering* Washington, D.C.: National Academies Press, 2020, ISBN: 978-0-309-49981-1. DOI: 10.17226/25620. Available: https://www.nap.edu/catalog/25620.

[264] H. Wang, A. Khajepour, D. Cao and T. Liu, “Ethical Decision Making in Autonomous Vehicles: Challenges and Research Progress,” *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 1, pp. 6–17, 2022, DOI: 10.1109/MITS.2019.2953556.

[265] B. M. McLaren, “Computational models of ethical reasoning: Challenges, initial steps, and future directions,” *Machine Ethics*, vol. 9780521112, no. August, pp. 297–315, 2011, ISBN: 9780511978036. DOI: 10.1017/CBO9780511978036.018.

[266] P. Robinson, L. Sun, H. Furey, R. Jenkins, C. R. Phillips, et al., “Modelling Ethical Algorithms in Autonomous Vehicles Using Crash Data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 7775–7784, 2022, DOI: 10.1109/TITS.2021.3072792. Available: https://doi.org/10.1109/TITS.2021.3072792,.

[267] E. Dogan, F. Costantini and R. Le Boennec, “Ethical issues concerning automated vehicles and their implications for transport,” in *Advances in Transport Policy and Planning*. vol. 5 Academic Press, 2020, pp. 215–233, ISBN: 9780128201916. DOI: 10.1016/bs.atpp.2020.05.003.

[268] M. Geisslinger, F. Poszler and M. Lienkamp, “An Ethical Trajectory Planning Algorithm for Autonomous Vehicles,” *Nature Machine Intelligence*, vol. 5, pp. 137–144, 2023, DOI: https://doi.org/10.1038/s42256-022-00607-z. Available: https://doi.org/10.1038/s42256-022-00607-z.

[269] M. Geisslinger and TUM - Institute of Automotive Technology. “*TUMFTM/EthicalTrajectoryPlanning: Initial Release*,” June 2022. DOI: 10.5281/ZENODO.6684625. Available: https://doi.org/10.5281/zenodo.6684625#.YtAthEW-RUs.mendeley.

[270] M. Geisslinger, R. Trauth, G. Kaljavesi and M. Lienkamp, “Maximum Acceptable Risk as Criterion for Decision-Making in Autonomous Vehicle Trajectory Planning,” *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 570–579, 2023, DOI: 10.1109/OJITS.2023.3298973. Available: https://ieeexplore.ieee.org/document/10195149/.

[271] M. Geisslinger, P. Karle, J. Betz and M. Lienkamp, “Watch-and-Learn-Net: Self-supervised Online Learning for Probabilistic Vehicle Trajectory Prediction,” in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 869–875, ISBN: 9781665442077. DOI: 10.1109/smc52423.2021.9659079.

[272] M. Werling, J. Ziegler, S. Kammel and S. Thrun, “Optimal trajectory generation for dynamic street scenarios in a frenét frame,” *Proceedings - IEEE International Conference on Robotics and Automation*, no. March 2015, pp. 987–993, 2010, ISBN: 9781424450381. DOI: 10.1109/ROBOT.2010.5509799.

[273] H. Winner, “Handbuch Fahrerassistenzsysteme,” in *Handbuch Fahrerassistenzsysteme* Wiesbaden: Springer Fachmedien Wiesbaden, 2015, ISBN: 9783658057336. DOI: 10.1007/978-3-658-05734-3_62. Available: http://link.springer.com/10.1007/978-3-658-05734-3_62.

[274] F. Pfab, “Motion Planning with Risk Assessment for Automated Vehicles,” Technical University of Munich, 2020.

[275] G. A. Holton, “Defining risk,” *Financial Analysts Journal*, vol. 60, no. 6, pp. 19–25, 2004, DOI: 10.2469/faj.v60.n6.2669.

[276] M. Geisslinger and TUM - Institute of Automotive Technology. “*Wale-Net Prediction Network for CommonRoad*,” 2021. Available: https://github.com/TUMFTM/Wale-Net.

[277] P. Bender, J. Ziegler and C. Stiller, “Lanelets: Efficient map representation for autonomous driving,” *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, pp. 420–425, 2014, ISBN: 9781479936380. DOI: 10.1109/IVS.2014.6856487.

[278] I. Sutskever, O. Vinyals and Q. V. Le, “Sequence to sequence learning with neural networks,” *Advances in Neural Information Processing Systems*, vol. 4, no. January, pp. 3104–3112, 2014.

[279] World Health Organization. “*WHOQOL: Measuring Quality of Life*,” Available: https://www.who.int/toolkits/whoqol.

[280] A. Sobhani, W. Young, D. Logan and S. Bahrololoom, “A kinetic energy model of two-vehicle crash injury severity,” *Accident Analysis and Prevention*, vol. 43, no. 3, pp. 741–754, 2011, DOI: 10.1016/j.aap.2010.10.021. Available: http://dx.doi.org/10.1016/j.aap.2010.10.021.

[281] E. Rosén and U. Sander, “Pedestrian fatality risk as a function of car impact speed,” *Accident Analysis and Prevention*, vol. 41, no. 3, pp. 536–542, 2009, DOI: 10.1016/j.aap.2009.02.002.

[282] National Highway Traffic Safety Administration, “Crash Report Sampling System,” Available: https://www.nhtsa.gov/crash-data-systems/crash-report-sampling-system.

[283] T. A. Gennarelli and E. Wodzin, “AIS 2005: A contemporary injury scale,” *Injury*, vol. 37, no. 12, pp. 1083–1091, 2006, DOI: 10.1016/j.injury.2006.07.009.

[284] T. Geißenberger, “Harm Prediction for Risk-Aware Motion Planning of Automated Vehicles,” Technical University of Munich, 2021.

[285] J. Hey, *Experimental Economics*, Physica-Verlag HD, 2013, ISBN: 9783642511790. Available: https://books.google.de/books?id=pyLzCAAAQBAJ.

[286] E. J. Emanuel, G. Persad, R. Upshur, B. Thome, M. Parker, et al., “Fair Allocation of Scarce Medical Resources in the Time of Covid-19,” *New England Journal of Medicine*, vol. 382, no. 21, pp. 2049–2055, 2020, DOI: 10.1056/nejmsb2005114.

[287] J. Savulescu, M. Vergano, L. Craxì and D. Wilkinson, “An ethical algorithm for rationing life-sustaining treatment during the COVID-19 pandemic,” *British Journal of Anaesthesia*, vol. 125, no. 3, pp. 253–258, 2020, DOI: 10.1016/j.bja.2020.05.028. Available: http://www.bjanaesthesia.org/article/S0007091220304104/fulltext%20http://www.bjanaesthesia.org/article/S0007091220304104/abstract%20https://www.bjanaesthesia.org/article/S0007-0912(20)30410-4/abstract.

[288] J. Nida-Rümelin, J. Schulenburg and B. Rath, *Risikoethik*, 2012, DOI: 10.1515/9783110219982.

[289] J. C. Harsanyi and B. J. C. Harsanyi, “Bayesian Decision Theory and Utilitarian Ethics,” *The American Economic Review*, vol. 68, no. 2, pp. 223–228, 1978. Available: https://www.jstor.org/stable/1816692.

[290] S. C. Wright and G. D. Boese, “Meritocracy and Tokenism,” in *International Encyclopedia of the Social & Behavioral Sciences* Elsevier, 2015, pp. 239–245, DOI: 10.1016/B978-0-08-097086-8.24074-9. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780080970868240749.

[291] D. Leben, *Ethics for Robots*, Routledge, 2018, ISBN: 9781315197128. DOI: 10.4324/9781315197128. Available: https://www.taylorfrancis.com/books/mono/10.4324/9781315197128/ethics-robots-derek-leben.

[292] C. Pek, S. Manzinger, M. Koschi and M. Althoff, “Using online verification to prevent autonomous vehicles from causing accidents,” *Nature Machine Intelligence*, vol. 2, no. 9, pp. 518–528, 2020, DOI: 10.1038/s42256-020-0225-y. Available: http://dx.doi.org/10.1038/s42256-020-0225-y.

[293] M. Althoff, “Reachability Analysis and its Application to the Safety Assessment of Autonomous Cars,” *Fakultät für Elektrotechnik und Informationstechnik*, p. 221, 2010, ISBN: 90-9019824-5. DOI: 10.1017/CBO9781107415324.004. Available: http://mediatum2.ub.tum.de/doc/963752/963752.pdf%5Cnhttp://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20100715-963752-1-4.

[294] M. Althoff, M. Koschi and S. Manzinger, “CommonRoad: Composable benchmarks for motion planning on roads,” *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 719–726, 2017, ISBN: 9781509048045. DOI: 10.1109/IVS.2017.7995802.

[295] S. Maierhofer, M. Klischat and M. Althoff, “CommonRoad Scenario Designer: An Open-Source Toolbox for Map Conversion and Scenario Creation for Autonomous Vehicles,” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2021-Septe, pp. 3176–3182, 2021, ISBN: 9781728191423. DOI: 10.1109/ITSC48978.2021.9564885.

[296] M. Klischat and M. Althoff, “Generating critical test scenarios for automated vehicles with evolutionary algorithms,” *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2019-June, no. Iv, pp. 2352–2358, 2019, ISBN: 9781728105604. DOI: 10.1109/IVS.2019.8814230.

[297] M. Klischat, E. I. Liu, F. Höltke and M. Althoff, “Scenario Factory: Creating Safety-Critical Traffic Scenarios for Automated Vehicles,” *2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020*, 2020, ISBN: 9781728141497. DOI: 10.1109/ITSC45102.2020.9294629.

[298] CommonRoad. “*Motion Planning Competition for Autonomous Vehicles 2023 - Phase 1*,” 2023. Available: https://commonroad.in.tum.de/challenges/id/a7cf5e8b-660d-4d56-bb53-7d352eb33a83.

[299] CommonRoad. “*Motion Planning Competition for Autonomous Vehicles 2023 - Phase 2*,” 2023. Available: https://commonroad.in.tum.de/challenges/id/d5a01f69-e828-436d-9f9d-dabf4226e29f.

[300] R. Trauth, P. Karle, T. Betz and J. Betz, “An End-to-End Optimization Framework for Autonomous Driving Software,” *2023 3rd International Conference on Computer, Control and Robotics (ICCCR)*, pp. 137–144, 2023, ISBN: 9781665492126. DOI: 10.1109/ICCCR56747.2023.10193889.

[301] P. Allison, *Multple Regression: A Primer*, London, Sage Publications, 1998, ISBN: 9780761985334.

[302] A. Moore, “Ethical Theory, Completeness & Consistency,” *Ethical Theory and Moral Practice*, vol. 10, no. 3, pp. 297–308, 2007, DOI: 10.1007/sl0677-007-9070-x.

[303] F. Poszler, M. Geisslinger, J. Betz and C. Lütge, “Applying ethical theories to the decision-making of self-driving vehicles: A systematic review and integration of the literature,” *Technology in Society*, vol. 75, p. 102350, 2023, DOI: 10.1016/j.techsoc.2023.102350. Available: https://linkinghub.elsevier.com/retrieve/pii/S0160791X23001550.

[304] A. Grinbaum, “Chance as a value for artificial intelligence,” *Journal of Responsible Innovation*, vol. 5, no. 3, pp. 353–360, 2018, DOI: 10.1080/23299460.2018.1495032. Available: https://doi.org/10.1080/23299460.2018.1495032.

[305] C. Lütge, F. Poszler, A. J. Acosta, D. Danks, G. Gottehrer, et al., “AI4people: Ethical guidelines for the automotive sector-fundamental requirements and practical recommendations,” *International Journal of Technoethics*, vol. 12, no. 1, pp. 101–125, 2021, DOI: 10.4018/IJT.20210101.oa2.

[306] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, et al., “An open approach to autonomous vehicles,” *IEEE Micro*, vol. 35, no. 6, pp. 60–68, 2015, DOI: 10.1109/MM.2015.133.

[307] S. Kato, S. Tokunaga, Y. Maruyama, S. Maeda, M. Hirabayashi, et al., “Autoware on Board: Enabling Autonomous Vehicles with Embedded Systems,” *Proceedings - 9th ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS 2018*, pp. 287–296, 2018, ISBN: 9781538653012. DOI: 10.1109/ICCPS.2018.00035.

[308] P. Karle, T. Betz, M. Bosk, F. Fent, N. Gehrke, et al., “EDGAR: An Autonomous Driving Research Platform – From Feature Development to Real-World Application,” *arXiv*, 2023. Available: http://arxiv.org/abs/2309.15492.

[309] F. Favaro, L. Fraade-Blanar, S. Schnelle, T. Victor, M. Peña, et al., *Building a Credible Case for Safety: Waymo’s Approach for the Determination of Absence of Unreasonable Risk*, 2023, ISBN: 2020010607.

[310] Cruise, “Cruise Safety Report 2022,” 2022. Available: https://assets.ctfassets.net/95kuvdv8zn1v/zKJHD7X22fNzpAJztpd5K/ac6cd2419f2665000e4eac3b7d16ad1c/Cruise_Safety_Report_2022_sm-optimized.pdf.

[311] S. Riedmaier, T. Ponn, D. Ludwig, B. Schick and F. Diermeyer, “Survey on Scenario-Based Safety Assessment of Automated Vehicles,” *IEEE Access*, vol. 8, pp. 87456–87477, 2020, DOI: 10.1109/ACCESS.2020.2993730.

Prior Publications

During the development of this dissertation, publications and student theses were written in which partial aspects of this work were presented.

Journals; Scopus/Web of Science listed (peer-reviewed)

[51] M. Geisslinger, F. Poszler, J. Betz, C. Lütge and M. Lienkamp, "Autonomous Driving Ethics: from Trolley Problem to Ethics of Risk," *Philosophy & Technology*, 2021, DOI: 10.1007/s13347-021-00449-4. Available: https://doi.org/10.1007/s13347-021-00449-4%20https://link.springer.com/10.1007/s13347-021-00449-4.

[152] J. Betz, T. Betz, F. Fent, M. Geisslinger, A. Heilmeier, et al., "TUM Autonomous Motorsport: An Autonomous Racing Software for the Indy Autonomous Challenge," *Journal of Field Robotics*, 2023, DOI: 10.1002/ROB.22153. Available: https://onlinelibrary.wiley.com/doi/10.1002/rob.22153%20http://arxiv.org/abs/2205.15979.

[228] P. Karle, M. Geisslinger, J. Betz and M. Lienkamp, "Scenario Understanding and Motion Prediction for Autonomous Vehicles - Review and Comparison," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 16962–16982, 2022, DOI: 10.1109/TITS.2022.3156011. Available: https://www.ieee.org/publications/rights/index.html.

[268] M. Geisslinger, F. Poszler and M. Lienkamp, "An Ethical Trajectory Planning Algorithm for Autonomous Vehicles," *Nature Machine Intelligence*, vol. 5, pp. 137–144, 2023, DOI: https://doi.org/10.1038/s42256-022-00607-z. Available: https://doi.org/10.1038/s42256-022-00607-z.

[270] M. Geisslinger, R. Trauth, G. Kaljavesi and M. Lienkamp, "Maximum Acceptable Risk as Criterion for Decision-Making in Autonomous Vehicle Trajectory Planning," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 570–579, 2023, DOI: 10.1109/OJITS.2023.3298973. Available: https://ieeexplore.ieee.org/document/10195149/.

[303] F. Poszler, M. Geisslinger, J. Betz and C. Lütge, "Applying ethical theories to the decision-making of self-driving vehicles: A systematic review and integration of the literature," *Technology in Society*, vol. 75, p. 102350, 2023, DOI: 10.1016/j.techsoc.2023.102350. Available: https://linkinghub.elsevier.com/retrieve/pii/S0160791X23001550.

F. Poszler, M. Geisslinger and C. Lütge, "Ethical decision-making for self-driving vehicles: A proposed model & checklist of underlying values that need to be concretized in the future," *Science and Engineering Ethics*, vol. under Revi, 2023.

Conferences, Periodicals; Scopus/Web of Science listed (peer-reviewed)

[174] R. Trauth, M. Kaufeld, M. Geisslinger and J. Betz, “Learning and Adapting Behavior of Autonomous Vehicles through Inverse Reinforcement Learning,” in *2023 IEEE Intelligent Vehicles Symposium (IV)*, 2023, pp. 1–8, DOI: 10.1109/IV55152.2023.10186668.

[207] A. Wischnewski, M. Geisslinger, J. Betz, T. Betz, F. Fent, et al., “Indy Autonomous Challenge - Autonomous Race Cars at the Handling Limits,” in *12th International Munich Chassis Symposium 2021* Springer Berlin Heidelberg, 2022, pp. 1–16.

[271] M. Geisslinger, P. Karle, J. Betz and M. Lienkamp, “Watch-and-Learn-Net: Self-supervised Online Learning for Probabilistic Vehicle Trajectory Prediction,” in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 869–875, ISBN: 9781665442077. DOI: 10.1109/smc52423.2021.9659079.

F. Poszler, M. Geisslinger and C. Lütge, “A five-step ethical decision-making model for self-driving vehicles: Which (ethical) theories could guide the process and what values need further investigation?,” in *International Conference on Computer Ethics*, 2023, pp. 1–3.

Journals, Conferences, Periodicals, Reports, Conference Proceedings and Poster, etc.; not Scopus/Web of Science listed

[308] P. Karle, T. Betz, M. Bosk, F. Fent, N. Gehrke, et al., “EDGAR: An Autonomous Driving Research Platform – From Feature Development to Real-World Application,” *arXiv*, 2023. Available: http://arxiv.org/abs/2309.15492.

F. Poszler and M. Geißlinger, “AI and Autonomous Driving : Key ethical considerations,” no. February, 2021.

Non-thesis-relevant publications; Scopus/Web of Science listed (peer-reviewed)

F. Nobis, M. Geisslinger, M. Weber, J. Betz and M. Lienkamp, “A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection,” in *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2019, pp. 1–7, ISBN: 978-1-7281-5085-7. DOI: 10.1109/SDF.2019.8916629. Available: https://ieeexplore.ieee.org/document/8916629/.

A. Heilmeier, M. Geisslinger and J. Betz, “A Quasi-Steady-State Lap Time Simulation for Electrified Race Cars,” in *2019 14th International Conference on Ecological Vehicles and Renewable Energies, EVER 2019*, 2019, pp. 1–10, ISBN: 9781728137032. DOI: 10.1109/ever.2019.8813646.

P. Karle, F. Török, M. Geisslinger and M. Lienkamp, “MixNet: Physics Constrained Deep Neural Motion Prediction for Autonomous Racing,” *IEEE Access*, vol. 11, no. August, pp. 85914–85926, 2023, DOI: 10.1109/ACCESS.2023.3303841. Available: http://arxiv.org/abs/2208.01862%20https://ieeexplore.ieee.org/document/10214014/.

Thesis-relevant open-source software

[269] M. Geisslinger and TUM - Institute of Automotive Technology. "*TUMFTM/EthicalTrajectoryPlanning: Initial Release*," June 2022. DOI: 10.5281/ZENODO.6684625. Available: https://doi.org/10.5281/zenodo.6684625#.YtAthEW-RUs.mendeley.

[276] M. Geisslinger and TUM - Institute of Automotive Technology. "*Wale-Net Prediction Network for CommonRoad*," 2021. Available: https://github.com/TUMFTM/Wale-Net.

Supervised Student Theses

The following student theses were written within the framework of the dissertation under the supervision of the author in terms of content, technical and scientific support as well as under relevant guidance of the author. In the following, the bachelor, semester and master theses relevant and related to this dissertation are listed. Many thanks to the authors of these theses for their extensive support within the framework of this research project.

[274] F. Pfab, “Motion Planning with Risk Assessment for Automated Vehicles,” Technical University of Munich, 2020.

[284] T. Geißenberger, “Harm Prediction for Risk-Aware Motion Planning of Automated Vehicles,” Technical University of Munich, 2021.

F. Girstl, “Autonomous Driving : Probability-based Vehicle Trajectory Prediction,” Technical University of Munich, 2020.

C. Krispler, “Motion Planning for Autonomous Vehicles: Developing a Principle of Responsibility for Ethical Decision-Making,” Technical University of Munich, 2022.

S. Sagmeister, “Neural Networks: Real-time Capable Trajectory Planning through Supervised Learning,” Technical University of Munich, 2021.

L. Bayerlein, “Scenario Geneartion for Assessment of Autonomous Driving Functionilities,” Technical Univeristy of Munich, 2021.

F. Lattemann, “Stochastic Optimization of Hyperparameters for Dynamic Trajectory Planning of Autonomous Vehicles,” Technical University of Munich, 2021.

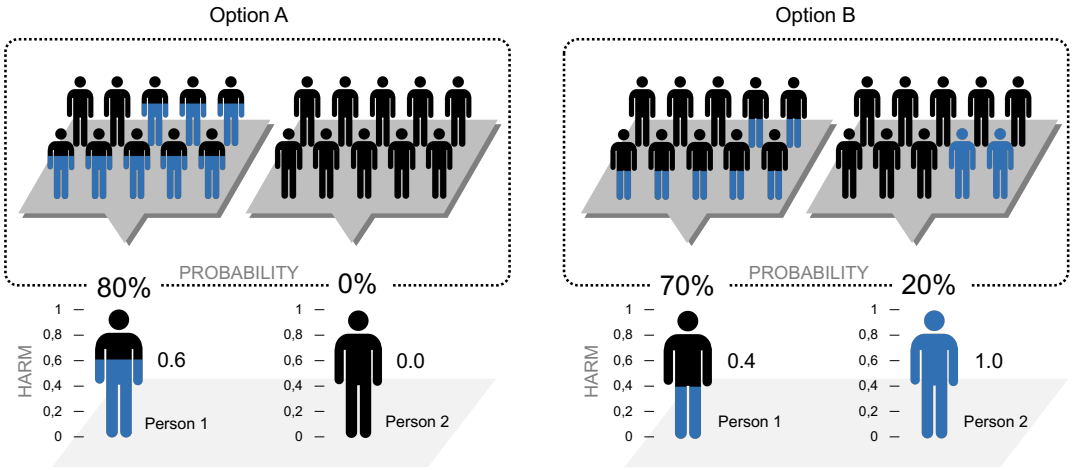
D. Scholz, “Trajectory Planning using Reinforcement Learning,” Technical University of Munich, 2021.

M. Müller, “Uncertainty Evaluation of Obscured Sensors for Autonomous Vehicles,” Technical University of Munich, 2021.

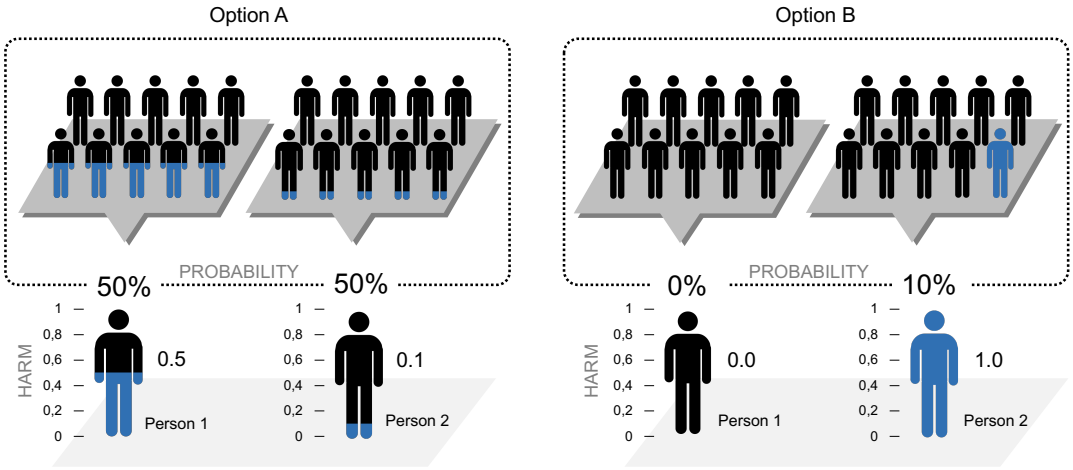
Appendix

A Questions of the Ethical Vehicle Experiment

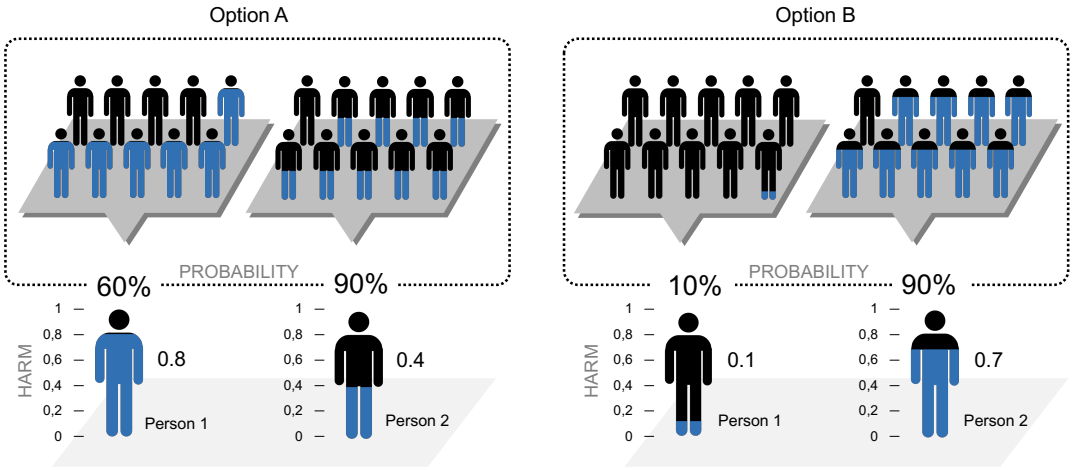
Question 1



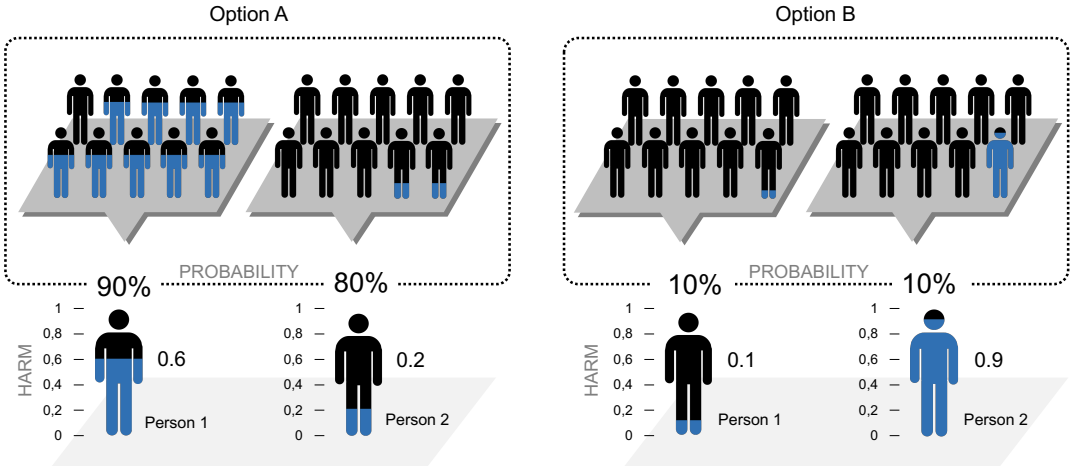
Question 2



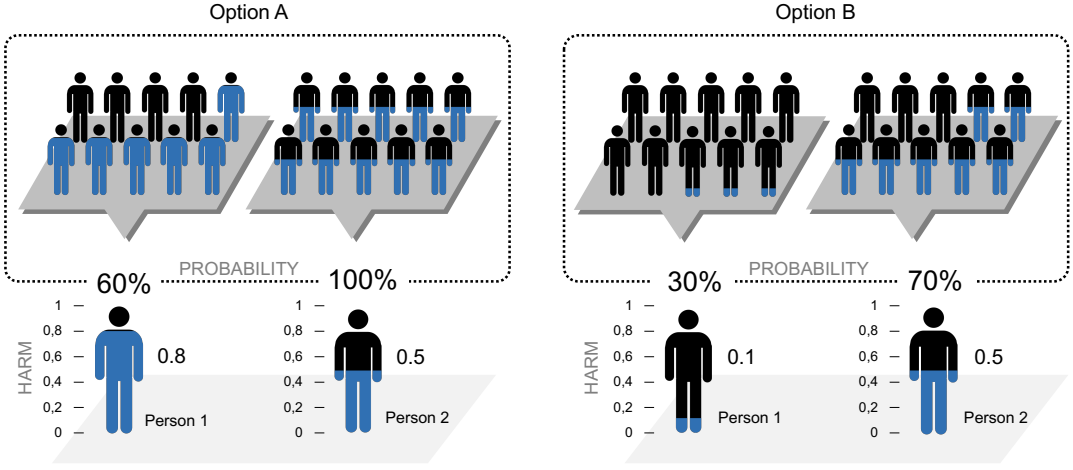
Question 3



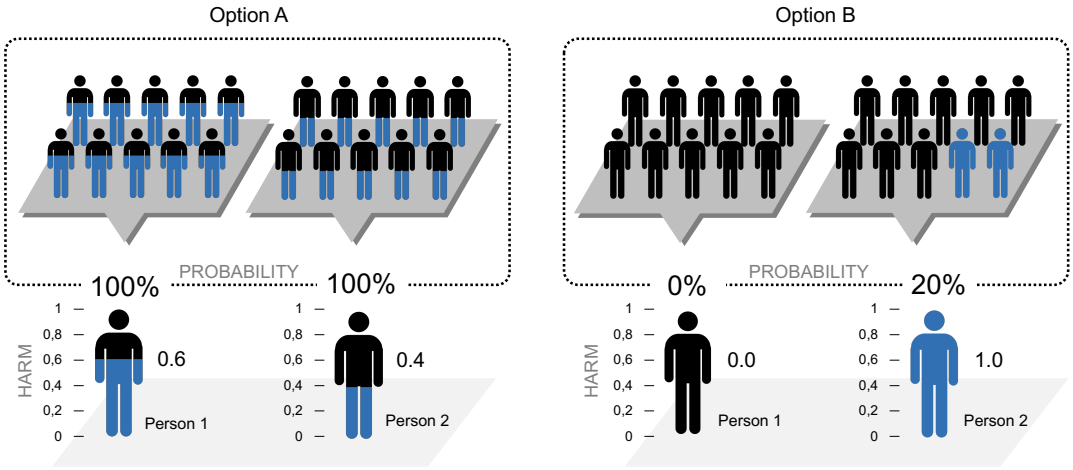
Question 4



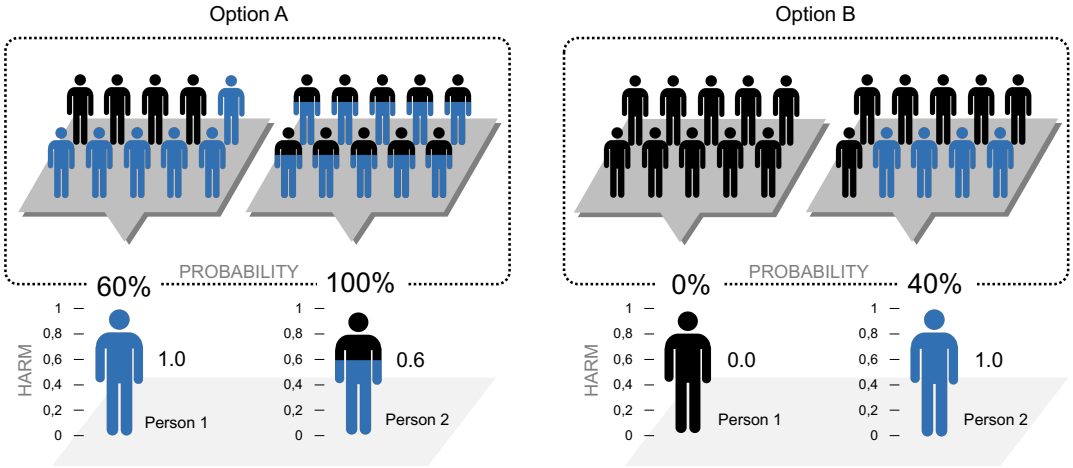
Question 5



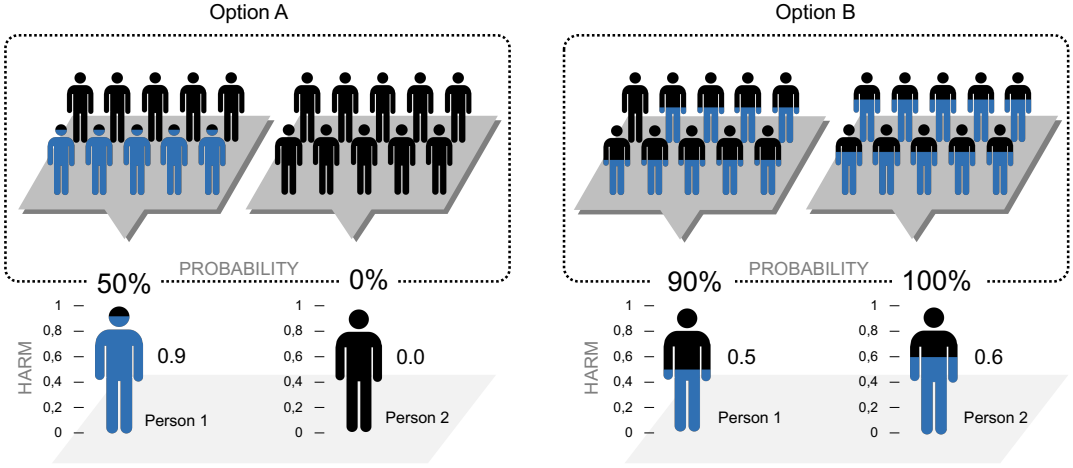
Question 6



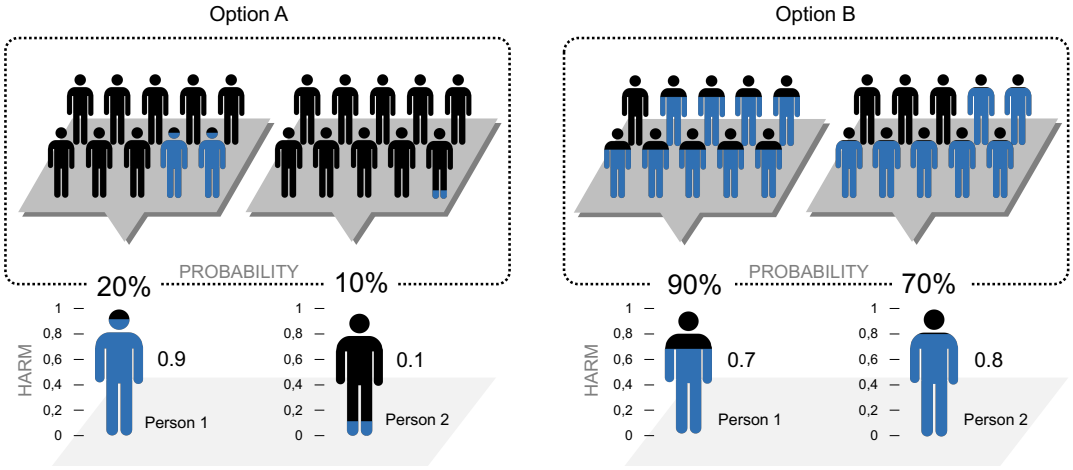
Question 7



Question 8



Question 9


Option A
PROBABILITY
20%
10%
HARM
1
0,8
0,6
0,4
0,2
0
0.9
0.1
Person 1
Person 2
Option B
PROBABILITY
90%
70%
HARM
0.7
0.8
Person 1
Person 2