Received 13 April 2023; revised 14 June 2023; accepted 22 July 2023. Date of publication 26 July 2023; date of current version 14 August 2023.

Digital Object Identifier 10.1109/OJITS.2023.3298973

Maximum Acceptable Risk as Criterion for Decision-Making in Autonomous Vehicle Trajectory Planning

MAXIMILIAN GEISSLINGER[1,2], RAINER TRAUTH[1,2], GEMB KALJAVESI[1,2], AND MARKUS LIENKAMP[1,2] (Member, IEEE)

[1]Institute of Automotive Technology, Technical University of Munich, 85748 Garching, Germany

[2]Munich Institute of Robotics and Machine Intelligence, Technical University of Munich, 85748 Garching, Germany

CORRESPONDING AUTHOR: M. GEISSLINGER (e-mail: maximilian.geisslinger@tum.de)

This work was supported in part by the Technical University of Munich-Institute for Ethics in Artificial Intelligence (IEAI); in part by the Bavarian Research Foundation (BFS); and in part by the Project Wies'n Shuttle in the MCube Cluster under Grant FKZ 03ZU1105AA.

ABSTRACT Autonomous vehicles are being developed to make road traffic safer in the future. The time when autonomous vehicles are actually safe enough to be used in real traffic is a current subject of discussion between industry, science, and society. In our work, we propose a new approach to the risk assessment of autonomous vehicles based on risk-benefit analysis, as it is already established in other areas, such as the registration of pharmaceuticals. In this context, we address the question of socially acceptable risk for mobility and investigate this concept as a decision-making criterion in trajectory planning. We make the first attempt to quantify an accepted risk by comparing autonomous vehicles with other types of mobility while taking into account the ethical and psychological effects important to the acceptance of autonomous vehicles. We show how an accepted risk contributes to the transparent decision-making of autonomous vehicles at the maneuver level. Finally, we present a method for considering accepted risk in trajectory planning. The evaluation of this algorithm in a simulation of 2,000 scenarios reveals that lower risk thresholds can actually reduce risks in trajectory planning. The code used in this research is publicly available as open-source software: https://github.com/TUMFTM/EthicalTrajectoryPlanning.

INDEX TERMS autonomous vehicles, trajectory planning, decision-making, risk.

I. INTRODUCTION

AUTONOMOUS Vehicles (AVs) will be subject to inherent risks which cannot be completely mitigated [1], [2], [3]. This raises questions about risk management, including what risks can be accepted. The following illustrative example should motivate our work and show the importance of risk for the behavioral decision-making of AVs.

Imagine the following situation: An autonomous vehicle is driving through the city. Two children are playing on the sidewalk at the side of the road. How should the AV pass by the playing children? Should it maintain its speed of 40 km/h and increase the safety distance by driving slightly into the seemingly clear adjacent lane? Or should it lower its speed to walking speed (7 km/h) for safety reasons?

Given this fictitious but realistic situation, three things become apparent:

1) No matter how safe AVs are and how well their software is programmed, there is always a residual risk for an accident. In almost every case, there is a possibility that one of the children will hop onto the road in such a way that the AV can no longer avoid a collision.
2) The information described in the situation is not sufficient to evaluate the situation in a meaningful way. What are, for example, the distances between the vehicle and the children? How likely is it that one of the

The review of this article was arranged by Associate Editor Fernando Auat Cheein.

children will run into the road? Evaluating the situation requires knowledge of risks (more on this later).

3) We see a trade-off between safety and traffic flow. Reducing risk in road traffic is generally achieved by reducing speed and increasing safety distances. Both, however, lead to a worse traffic flow. After all, what would be the benefit of AVs if they were significantly safer but were not used because of their slow and overly defensive driving behavior? If AVs drastically worsen traffic congestion in already stressed cities because of their driving behavior, this may lead to similar acceptance problems as dubious safety.

Ultimately, this trade-off raises a crucial underlying question of how much risk we want to accept in autonomous driving. The answer to this question also provides the answer to the aforementioned situation. If we know a value for acceptable risk, the vehicle can choose its speed, or more generally its driving behavior, in such a way that this risk is not exceeded. A society-accepted risk could consequently be the basis for the decision-making of AVs. Whether or not to overtake a slow cyclist on a country road, for example, depends on many variables, all of which ultimately result in a risk.

We will see that a weighing of risks is not new at all. Several fields, such as the registration of pharmaceuticals, apply methods of risk-benefit analysis to decisions that involve uncertainty. For this reason, various calculations concerning road traffic attributed a monetary value to human life, such as 9.6 million dollars for example by the U.S. department of transportation [4], in order to be able to systematically weigh up political decisions that are related to risks for humans.

This contribution seeks to develop a procedure used to quantify a socially accepted risk and to consider it accordingly in the trajectory planning of AVs. In doing so, we not only want to illuminate the perspective of the potential users of AVs, but also focus on the other road users who are directly affected by the behavior of AVs. Studies have shown that the interests of AV users may well lead to an ethical conflict with those of other road users [5]. For this reason, i.e., in order to represent the interests of all road users, the search for an acceptable risk must be conducted as a social debate.

II. RELATED WORKS

The literature relevant to the present work comprises four areas: First, in terms of risk management, the safety standards and norms are briefly presented in order to later connect risks and safety to trajectory planning as part of the AV decision-making process. Therefore, approaches from formal verification are considered next to risk-aware trajectory planning. Finally, a look at the risk-benefit analysis in the pharmaceutical sector is intended to provide new inspiration and ideas for AV risk management.

A. SAFETY STANDARDS AND LEGISLATION IN AUTONOMOUS DRIVING

AV risk and safety issues are currently addressed by standards such as ISO 26262 [6] or ISO/PAS 21448 SOTIF [7]. While ISO 26262 aims to ensure the functional safety of electric and electronic systems in an (automated) vehicle, SOTIF focuses on the safety of the intended functionality. ISO 26262 uses the Automotive Safety Integrity Level (ASIL) as the unit of measurement. These ASIL levels build up on the SIL levels of ISO and thus allow us to connect a qualitative measurement from 10^{-5} to 10^{-9} dangerous failures per hour, although they are usually of qualitative nature [6]. The ASIL is determined by considering severity, probability of exposure, and controllability. It determines which safety measures must be deployed so that only an acceptable residual risk remains. The consideration of risk in ISO 26262 is based on individual components and the causes of the probability of failure. ISO 26262 acknowledges that safety is not the absence of risk, but rather the idea that some inherent risk will remain in the system [8]. However, there is no consideration of risk based on the consequences at the overall system level, and it is not designed to cope with multi-agent environments [9]. As a result, this standard cannot be connected to trajectory planning and decision-making of AVs, nor can it be used to answer the questions raised in the introduction. The system-level view is provided by SOTIF, which distinguishes four categories of risk: known safe, unknown safe, known unsafe, and unknown unsafe. The classification of risk into four different scenario types may give first indications about the question of acceptable risk, but it does not answer this question to a sufficient degree. The recent EU regulation on autonomous driving systems (ADS) takes the next step by demanding and further describing the usage of acceptance criteria in terms of residual risks [10]: Considering the ODD, "the manufacturer shall define acceptance criteria from which the validation targets of the ADS are derived to evaluate the residual risk" [10] based on various existing data, such as accident data. Based on accident data in the EU, the regulation provides an exemplary value of 10^{-7} fatalities per hour as an acceptance criterion in a footnote.

B. FORMAL VERIFICATION AND RESPONSIBILITY

Formal verification methods provide a connection between the safety and trajectory planning of AVs. The underlying idea is to prove safety in a planned trajectory based on a mathematical model for safety assurance [11]. Therefore, a safe driving policy for trajectory planning is often described using time-variant areas that the AV must avoid. Accordingly, reachable sets of road users [12] are calculated based on their current state and the physical and legal constraints of each road user. However, these sets grow exponentially over time in their size, which makes their use infeasible for longer planning horizons [13]. A similar approach is provided by the Responsibility Safety Shield (RSS) [9], which was developed with the specific goal of not leading to accidents in which the AV is at fault. However, choosing conservative RSS parameters likely hinders traffic flow, whereas the opposite could lead to collisions because the assumptions are violated [14]. Similar to the RSS approach, a reachability-based trajectory design is introduced in [15], which generates provable not-at-fault trajectories. Initial criticism based on accident data coming from the first AVs in the United States appears, stating that AVs may rather avoid blame than accidents [16]. This indicates that such not-at-fault approaches, or approaches that primarily aim to avoid only accidents resulting from the behavior of the AV, can increase the number of accidents overall and thus reduce road safety. 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 [17].

C. RISK-AWARE TRAJECTORY PLANNING

In contrast to guaranteeing safety based on the legal responsibilities of road users, risk-aware planning approaches acknowledge inherent risks in road traffic and aim to quantify and minimize these risks. Since the completed risk assessment during development cannot identify all potential hazardous events during real-time AV operations, there is an increasing interest in risk assessment during run time [18]. To consider risks in trajectory planning, criticality measures, such as time-to-collision (TTC) or time-to-react (TTR) are used in various approaches [19], [20]. Further risk-aware architectures used for AVs incorporate algorithmic uncertainties regarding various software functions, e.g., perception, intention detection, or control [21]. Building upon uncertainties, risk measures can be extended by integrating the severity of potential collisions [22], [23]. Experiments in a real-world application using Gaussian mixture models for modeling the uncertain behavior of obstacles exhibited benefits in terms of less conservative trajectories being able to be generated [24]. Further examples that focus on risks originating from sensor occlusion show great potential for improving AV safety and comfort [25].

Most of these risk-aware trajectory planning methods aim to minimize risks or uncertainties of the ego-AV. However, the requirements for AV risk assessment from an ethical point of view are more far-reaching and require the inclusion of the risks of other traffic participants [5]. In addition, no research has yet connected the idea of risk thresholds coming from a macroscopic view, e.g., crash statistics (lagging measures), using trajectory planning as a policy (using leading measures) for decision-making.

D. RISK-BENEFIT ANALYSIS

The EU Expert Group and the German Ethics Committee demand a positive risk balance for AVs, as compared to the human driver [26], [27]. This approach suggests a new method of safety argumentation that is already widely used in other areas, such as pharmaceuticals: a risk-benefit analysis.

According to risk-benefit analysis, risks and benefits are quantified and their ratio is applied as the basis for decision-making. Regarding the quantification, we distinguish leading and lagging measures [28], whereby leading measures are pre-crash measures and lagging measures reflect the post-crash outcomes. Relying on these lagging measures using crash statics for statistical safety validation requires hundreds of millions of kilometers [29] of real-world driving data for each software version, which motivates the investigation of leading measures as valid indicators for lagging measures.

Working with risk-benefit ratios yields the question of which ratio is considered acceptable when introducing AVs. Although the ethics committees suggest the human driver as a benchmark, it makes sense to look into other areas where risks must be weighed against benefits. It has been shown that the pharmaceutical and automobile sectors both share quite similar requirements and face similar magnitudes of both uncertainties and risks [30]. Therefore, initial work has begun on adapting the qualitative framework PrOACT-URL [31] from the pharmaceutical sector and applying it to the positive risk balance of AVs [30]. There are further relevant aspects that can be transferred to AVs: In addition to managing risks, it is also essential to educate users about existing risks in an accessible and understandable way. The patient leaflets that come with pharmaceuticals provide a valuable template for such a purpose. Another aspect concerns the difficulty of collecting sufficient data in advance in order to make valid apriori assumptions about the risks. Therefore, the risk-benefit ratio is continuously reviewed during use in the pharmaceutical sector. If the ratio becomes unfavorable, the drug should be withdrawn from the market.

III. HUMAN DRIVER AS A FIRST BENCHMARK

To provide an initial benchmark for accepted risk, we start by quantifying the risk that corresponds to the average human driver. Therefore, we have to clarify the definition of risk, because a variety of definitions are used in the literature, such as fatalities per hour [10] or likelihood of an accident [32] or collision [30]. Most risk definitions used in the context of AVs lack the two-dimensional character of risk by focusing on, e.g., fatalities or incidents alone, but neglecting the severity of an incident. Therefore, we selected the definition of risk as being an expected value consisting of the likelihood that an event occurs and a measure for the consequences of this event. Transferred to autonomous driving, we use the probability for a collision $p_{\text{Collision}}$ as likelihood and the expected harm H to humans as a consequence to calculate the risk R as in Equation 1. This equation has been established in previous contributions on the subject of risk-aware trajectory planning [5].

$$R = p_{\text{Collision}} \cdot H \tag{1}$$

In line with the risk-benefit analysis described in Section II-D, we do not set risk in contrast to time (risk per hour) as in ISO 26262 [6]. Following the idea of considering the benefits next to the risk, we compare the risk to

TABLE 1. Transfer of accidents statistics of 2021 [34] to a cumulative value for harm.

Accident Severity	Number	Harm per accident	Total harm per accident severity
Fatalities	2,562	1.0	256.2
Major injuries	55,137	0.1	5,513.7
Minor injuries	267,992	0.01	2,679.9
Total harm			**8,449.8**

the driven kilometers (risk per km), as this more effectively reflects the benefit of mobility.

Using Germany as an example, we converted data from accident statistics into a risk-per-kilometer-driven metric in order to create an initial benchmark for the average human driver. Whereas the risk factor considers any personal injury reported by the statistics, property damage and secondary effects, e.g., health issues due to air pollution, are not considered. Table 1 shows the number of fatalities, as well as major and minor injuries in German road traffic for 2021. In order to convert injuries and accident fatalities into risk, we rate the severity of an accident between 0 and 1, where 1 represents the worst accident injury, and 0 represents no injuries. In doing so, we rely upon the definition provided in [5] and map a value of 1.0 to death, 0.1 to major injuries, and 0.01 to minor injuries. This assignment is rather arbitrary in nature and remains open for discussion. However, the congruence with the harm model in the algorithm is important, so we do not address this issue in greater detail. Finally, this approach results in a total harm of $8,449.8$ through road traffic in Germany per year. Given the annual mileage of approximately $719.6 \cdot 10^9$ km in 2019 [33], we calculate a risk according to the average human driver of $\overline{R}_{\text{Human}} = 1.17 \cdot 10^{-8}$ per km.

This value reflects the risk of an average human driver in Germany. AVs at level 4 operate in various Operational Design Domains (ODD). As a result, this value cannot simply be applied to AVs. Essentially, the following aspects must also be taken into account:

- Risk is dependent on the ODD. For example, driving in good weather is inherently less risky than driving on slippery roads or in poor visibility. Accordingly, comparing AVs with human driving requires comparing human-driven accident data from the same ODD. The value provided herein reflects the average risk of a human driver in the ODD throughout Germany during the entire year.
- From an ethical point of view, it is not sufficient if AVs improve the average risk as opposed to human drivers. Focusing on the average risk only could result in higher risks for vulnerable road users (VRUs), for example, while the risks for the passengers in AVs decrease. This question raises an important ethical issue that requires looking at the balance of risk for different groups of road users. In particular, road users, who cannot choose between AVs or human-driven vehicles but only encounter the consequences (e.g., pedestrians), must then be granted a more positive risk balance compared to human-driven traffic.
- We must also consider the psychological effects influencing AV acceptance. Firstly, our benchmark incorporates the average human driver. As a result, some drivers may not see an improvement in risk from using AVs. In addition, studies have shown that human drivers overestimate their capabilities as a result of the overconfidence bias [35]. Thus, in the eyes of most drivers, the average human driving capability represents no improvement, even if it would actually be an improvement in many cases. Secondly, individuals perceive risks differently, depending on their ability to control risk. When having no control over the AV, individuals require the risk to be much smaller than for the same situation under their perceived control [36]. One survey among potential AV passengers from China revealed a risk improvement on the order of two magnitudes would be necessary for AVs to be broadly accepted [37].

Comparison with aviation or railway systems reveals that risks per km are roughly 100 times smaller in the case of trains and 1000 smaller in the case of planes [38]. This statistic suggests that a risk reduction by a factor of 100 in the corresponding ODD, and for all road users, represents a reasonable starting point for AV introduction leading to an average desired risk in the magnitude of 10^{-10} per km. Similar to pharmaceuticals, assessing the benefit-risk ratio at the time of registration can only be preliminary in nature. Drug studies demonstrate the lack of generalizability of the benefit-risk ratio derived from the pivotal studies [39]. Therefore, the risk-benefit analysis must be continuously reviewed with respect to the current state of knowledge. Given the situation where the ratio has become unfavorable (similar to the pharmaceutical industry), the vehicle registration might become invalid, which is a practice followed by most countries in terms of hardware issues (e.g., TÜV in Germany). This representation of risk is also suitable for communication with passengers. As is similar to patient information leaflets, problems which occur from frequently to rarely (e.g., injuries) can be presented to users in a transparent manner.

IV. TRAJECTORY PLANNING WITH MAXIMUM ACCEPTABLE RISK

To integrate the concept of a maximum acceptable risk criterion in AV trajectory planning, we build up on a sampling-based approach to trajectory planning [40]. In general, the planning algorithm consists of four steps, as shown by Figure 1: (1) generating jerk-optimal trajectories using a Frenet coordinate system, (2) identifying and discarding invalid trajectories, and (3) evaluating the valid trajectories according to a cost function and selecting the trajectory with the lowest cost (4). To integrate consideration of risk into AV decision-making, the trajectories generated as various

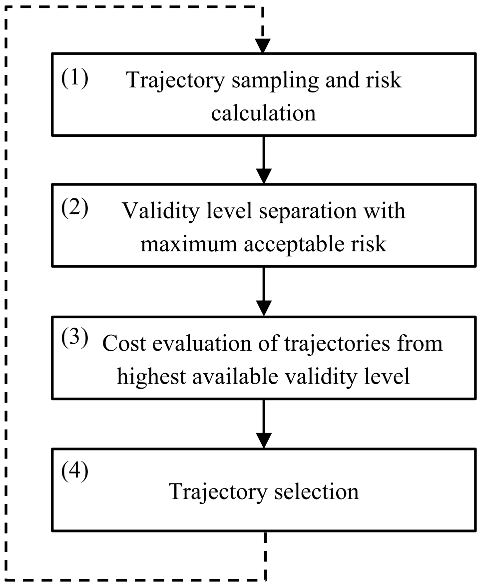


FIGURE 1. Schematic overview of the trajectory planning algorithm with maximum acceptable risk. The four steps are repeated with a frequency of 10 Hz.

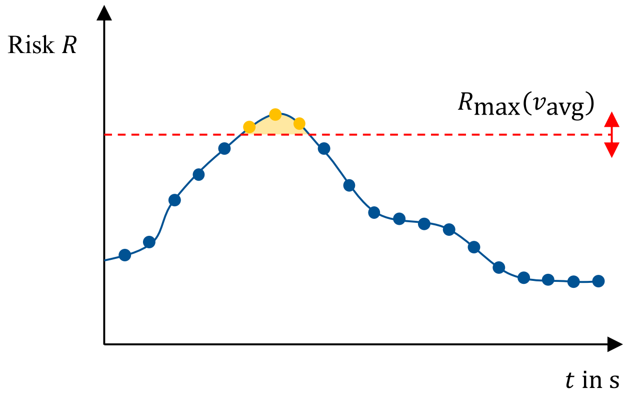


FIGURE 2. Exemplary risk over time for a sampled trajectory for a specific road user. The crucial criterion for evaluating the trajectory is the maximum risk that occurs over time.

decision options must each be assigned a risk value. There are various uncertainties in autonomous driving, such as perception, prediction, or control uncertainties, that can be transferred to collision probabilities. We use the probabilistic trajectory prediction model *Wale-Net* [41] to calculate collision probabilities for each state in every trajectory that was generated. The resulting overall collision probability from multiple possible collisions at a given state originating from several road users is calculated as conditional probability. Interactions between two or more predictions and the planned trajectory are represented by the data-based prediction model through training data. Further uncertainties, for example, originating from perception, are neglected here.

To estimate the harm of a potential collision, we use four inputs that have known physical relationships to the harm: velocities and masses of the two colliding objects, as well as impact areas and angles [42]. The correlations are expressed by a logistic regression whose parameters were determined based on the accident database of the National Highway Traffic Safety Administration's Crash Report Sampling System [43]. In order to meet the ethical requirements of prioritizing personal injury over property damage [44], only personal injury is considered by the model.

Finally, according to Equation (1), a risk for each state over time within the planning horizon of the AV can be calculated as shown by way of example in Figure 2. This results in a time-variant risk over the planning horizon for each sampled trajectory. We describe the risk for a trajectory with the maximum value over time since the correlations of the risk values (resulting from the uncertainties of the prediction model) are unknown. The underlying assumption of zero dependent risks results in a conservative estimate of the risk. However, as long as the dependencies are unknown, this seems to be the only reasonable option. A trajectory is classified as valid if its risk for every road user is less than the maximum accepted risk.

The set of all valid trajectories is calculated as follows: Let T be the set of all trajectories generated by the trajectory planner in the sampling step (1) that are to be checked for validity (2). We are looking for the set T_V of all valid trajectories, from which a trajectory is subsequently selected (3), for example, by using a cost function. By applying the maximum acceptable risk $R_{\max}$, the set T_V is obtained by Equation 2:

$$T_V = \begin{cases} \{\tau \in T \mid R_\tau \leq R_{\max}\} & \exists\, \tau \in T : R_\tau \leq R_{\max} \\ T & \text{otherwise} \end{cases} \tag{2}$$

A major task of AV trajectory planning is to balance various objectives. Various cost terms that are used in the literature [45] can be related to either safety goals, comfort goals, or mobility/efficiency goals. If no trajectory is found fulfilling the condition of maximum accepted risk, then the vehicle is in a high-risk situation. We then propose to evaluate the trajectories using the cost function 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. Our implementation neglects all costs other than risk if the maximum acceptable risk is exceeded (Equation (3)). The effects of this two-stage cost function depending on the maximum acceptable risk in contrast to a usual risk-based cost function and a cost function without risk will be part of our empirical evaluation.

$$J_{\text{total}} = \begin{cases} w_R J_{\text{Risk}} \\ \quad + w_M J_{\text{Mobility}} \\ \quad + w_C J_{\text{Comfort}} & \exists\, \tau \in T : R_\tau \leq R_{\max} \\ J_{\text{Risk}} & \text{otherwise} \end{cases} \tag{3}$$

Related works from the literature on risk-aware trajectory planning (see Section II-C) mainly use a cost function to integrate risk or uncertainties in trajectory planning. This maps a (primarily linear) trade-off between risk and the

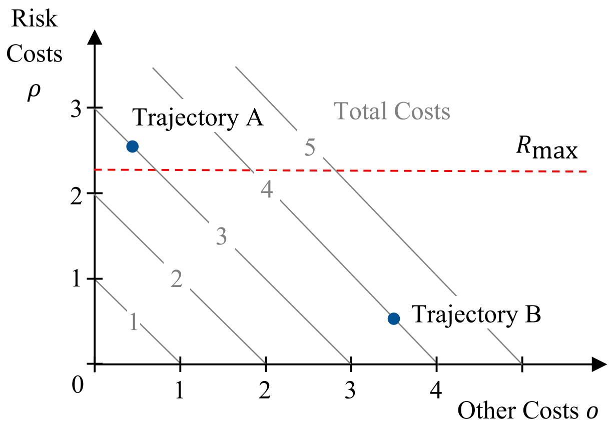


FIGURE 3. Relationship between maximum acceptable risk R_{max} and risk costs ρ assuming linear weights.

remaining costs. Thus, the risk is allowed to increase arbitrarily as long as this occurs within the ratio described by the cost function. Figure 3 illustrates a cost function with linear weights, distinguishing risk costs, and all the other costs. According to the cost function, the trajectory with the lowest cost must be chosen. In the visualized example, trajectory A must be preferred over B, although the risk costs are the highest. However, integrating the maximum acceptable risk would, in this case, exclude trajectory A given its excessive risk, so trajectory B becomes the better choice.

V. RESULTS

Our approach to trajectory planning with risk incorporates the concept of a maximum acceptable risk within the planning algorithm. In the following, we will examine the effects of this approach on trajectory planning and, in particular, the risks that arise as a result. Besides qualitative examples to illustrate the planner's behavior, we perform a quantitative evaluation with the CommonRoad [46] scenario library, which consists of 2,000 scenarios in the simulation. The scenarios originate from various ODDs with multiple countries (e.g., USA, China and Germany) and a wide variety of situations (e.g., city or highway). The behavior of various road users is deterministic, and thus they do not react according to the AV. In the simulation, a basic point mass model of the vehicle is employed, incorporating restrictions on acceleration and steering angles. The primary emphasis is placed on trajectory planning, assuming no variations in control between the planned trajectory and the executed trajectory. The scenarios represent real recorded scenarios and hand-crafted critical scenarios, our planning algorithm cannot achieve risks as low as demanded in Section III, and we stick to values that are more suitable here. Finally, we investigate the relationship between the online maximum acceptable risk as a leading measure in trajectory planning and the actual resulting risks as a lagging measure in road traffic.

A. QUALITATIVE EXAMPLES

In the introduction, we hypothesized that a maximum acceptable risk provides guidance to trajectory planning in terms of behavior and decision-making. The core of the proposed idea is to no longer prescribe AV behavior and evaluate corresponding metrics for safety or risk but to specify an accepted risk and derive a corresponding behavior of the AV from it. Accordingly, using a risk threshold is intended to set a suitable velocity in the introductory example or to specify whether or not to perform an overtaking maneuver. Since the latter is more illustrative, we use this example to test our hypothesis on a scenario with a potential overtaking maneuver. In our scenario, the white AV following our trajectory planning algorithm approaches a slow-moving scooter rider on a rural road in heavy oncoming traffic (represented in the form of trucks). The velocity of the scooter is significantly lower than the AV target speed here, which provokes an overtaking maneuver here.

The left of Figure 4 depicts the initial situation, without maximum acceptable risk (a) and $R_{max} = 10^{-7}$ (b). As indicated by the colors, the sampled trajectories that would overtake the scooter have lower costs (green) than those that do not overtake. This is mainly due to the high difference between the high AV target speed and the low speed in the case of not overtaking. The trade-off between safety and efficiency, as described by the cost function, advocates for overtaking. The lower the current velocity of the AV is, the more risk the AV is willing to take in that case. Consequently, the algorithm without maximum acceptable risks plans to overtake the scooter, as shown at t_1 and t_2.

Using the same planning algorithm with the same parameters, but with $R_{max} = 10^{-7}$ leads to different behavior. The low-cost trajectories aiming for the overtaking maneuver are then considered invalid because they exceed the maximum acceptable risk. As a result, the AV does not perform an overtaking maneuver and maintains a safe distance behind the scooter. The example demonstrates that a maximum accepted risk is appropriate for maneuver-level risk-based decisions. However, further experiments also show that the length of the planning horizon is a crucial parameter in this case: if the planning horizon is too short (e.g., 1 s instead of 2 s in this case), the risks that correspond to an overtaking maneuver can not be mapped to a trajectory choice so that the AV will enter a situation where it could not estimate the risk beforehand.

B. EMPIRICAL EVALUATION

The effects of the concept of maximum accepted risk described in the foregoing example impact the behavior in a variety of scenarios which can be evaluated from an empirical point of view. Therefore, we run our trajectory planning algorithm on 2,000 scenarios in simulation and compare various values for R_{max}. As a measure for comparison, we observe the resulting risk distribution across all of the scenarios. For this purpose, we analyze the risks of all road

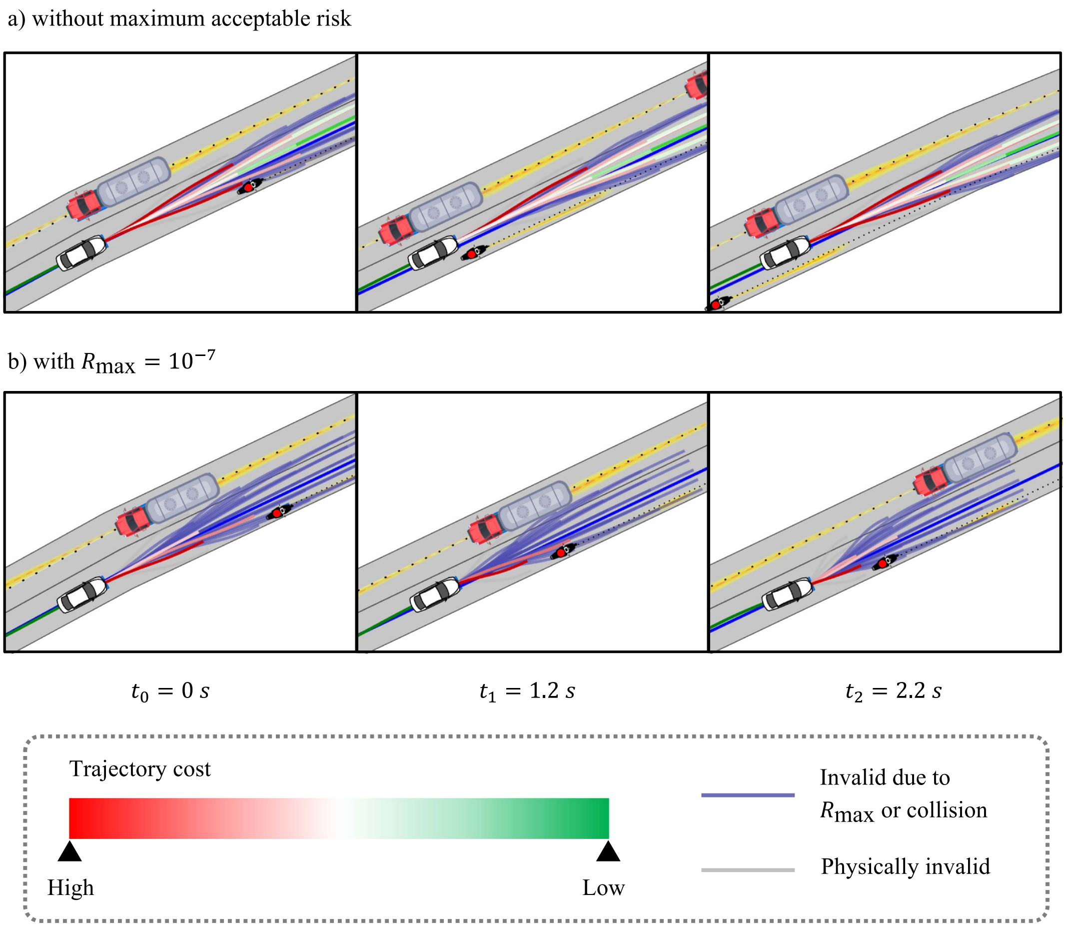


FIGURE 4. Overtaking maneuver as a qualitative example on the effects of a maximum acceptable risk in trajectory planning. The maneuver is performed with two algorithmic configurations shown at various timesteps: Without maximum acceptable risk (a), the white AV overtakes the slower scooter. Adding $R_{\max} = 10^{-7}$ to the same algorithm, an overtaking maneuver is considered unsafe, so the AV stays behind the scooter (b).

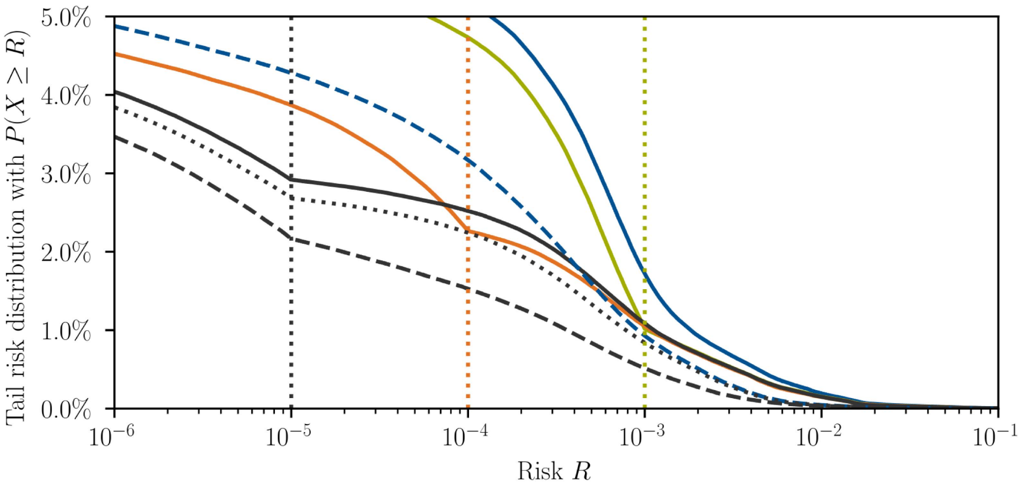
users collectively, which occur in each simulation time step with the respective configuration.

Figure 5 shows a tail distribution of these actual calculated risks for multiple settings. Reducing $R_{\max}$ from ∞ stepwise to 10^{-5} results in characteristic discontinuities at the relevant value of $R_{\max}$. This is indeed reasonable since the decision based on the maximum acceptable risk is not of a continuous nature but has a discrete character.

We also observe that lowering $R_{\max}$ does not necessarily decrease the average risk R_{mean} that can be actually measured in the simulated scenarios: Comparison between $R_{\max} = 10^{-4}$ and $R_{\max} = 10^{-5}$ reveals that although with $R_{\max} = 10^{-5}$ risks higher than 10^{-5} occur less frequently (3% vs. 4%) the average risk, in that case, is higher than with $R_{\max} = 10^{-4}$ ($7.25 \cdot 10^{-5}$ vs. $7.13 \cdot 10^{-5}$). This is true because higher risks (10^{-4}) are, in this case, more likely to occur than with $R_{\max} = 10^{-4}$.

We also investigate the impact of a risk-minimizing cost function in contrast and interaction with $R_{\max}$, as described in Section IV. First, we observe the AV behavior when only minimizing risk using the cost function without $R_{\max}$: Similar to the use of maximum acceptable risk, a risk-minimizing cost function decreases high risks but without the characteristics at the points of $R_{\max}$ (blue dashed line in Fig. 5). Combining both approaches results in even lower risks (black dotted line in Fig. 5): Compared to using $R_{\max} = 10^{-5}$ only, the average risk decreases by about 48% and, in comparison to using risk minimization only, by about 12%. However, the best results in terms of overall risk reduction are achieved by using the two-stage risk cost function described in Section IV: as soon as the maximum acceptable risk is exceeded, the objectives of comfort and mobility are neglected, and the planning algorithm minimizes only risk. This yields a further improvement of 48% (black dashed line) over the basic combination (blue dashed line).

An interesting effect appears when $R_{\max}$ is varied while using a risk-minimizing cost function as soon as $R_{\max}$ is exceeded. In this case, the average risk decreases as $R_{\max}$ is lowered, as Figure 6 illustrates. In contrast to Figure 6, where the lines cut each other, a basic correlation between $R_{\max}$ and the actual occurring risks can be observed. In addition, the limits of risk management emerge. The shaded area of the tail distribution shown in Figure 6 cannot be surpassed using the maximum risk and risk minimization



Line	R_{max}	Cost Function	R_{mean}
	∞	No risk included	$1.08 \cdot 10^{-4}$
	10^{-3}	No risk included	$8.29 \cdot 10^{-5}$
	10^{-4}	No risk included	$7.13 \cdot 10^{-5}$
	10^{-5}	No risk included	$7.25 \cdot 10^{-5}$
	∞	Incl. total risk minimization	$4{,}27 \cdot 10^{-5}$
	10^{-5}	Incl. total risk minimization	$3.74 \cdot 10^{-5}$
	10^{-5}	Incl. total risk minimization in 2 stages	$2.52 \cdot 10^{-5}$

FIGURE 5. **Tail distribution of actual risks occurring during simulation with 2000 CommonRoad scenarios with various values for maximum acceptable risks and three different cost functions.**

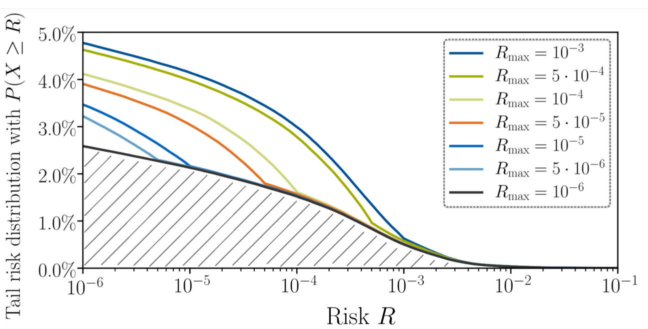


FIGURE 6. **Tail distribution of actual risks occurring during simulation with 2000 CommonRoad scenarios with varying values for R_{max} and a two-stage cost function with risk minimization.**

methods presented herein. Doing so would require adapting the AV algorithms. It should be noted that this limit in risk mitigation only emerges in simulations where the AV starts in the middle of a scenario with a prescribed state ($v > 0$). In a real-world application lowering the risk thresholds would rather lead to not starting a drive. This effect, however, offers promising opportunities for the benchmarking of trajectory planning algorithms within a simulation.

The correlations between the maximum accepted risk in trajectory planning and the actual resulting risks, as well as the resulting accidents, are of further interest. The ability to identify universally valid correlations in this context would open up new possibilities for the safety argumentation of AVs. Figure 7 illustrates this correlation with respect to two cases, i.e., no risk minimizing cost function and the two-stage risk minimization. As a result, we observe the average risk and cumulative harm over 2,000 scenarios. Compared to the case of no risk minimization, the two-stage cost function shows a correlation of maximum acceptable risk with the resulting average risk and harm: the lower R_{max}, the less cumulative harm and the less average risk. Beyond that, however, no additionally detailed relationships can be determined on the basis of these experiments. However, establishing a validated relationship between risks in trajectory planning and cumulative harm outputs of AVs based on algorithmic uncertainties might be able to open new avenues into the subject of safety validation. Assuming valid risk quantification, changes to a single software function would then no longer require repeated proof of the entire AV software package but rather only re-evaluation of the (expected) cumulative harm in the corresponding ODD.

VI. CONCLUSION & OUTLOOK

A wide variety of approaches have been proposed for controlling or eliminating autonomous driving risks. The present research addresses the question of how safe AVs must be in the context of risk-benefit analysis. We thereby found essential aspects from other fields (e.g., the pharmaceutical industry) to be appropriate when applied to the field of autonomous driving. In this context, we introduce the concept

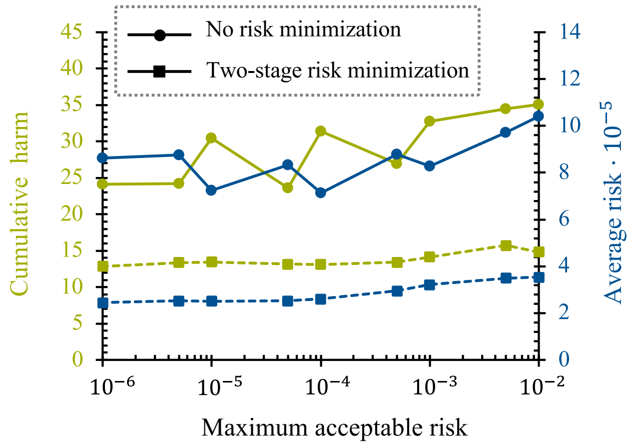


FIGURE 7. Relation between maximum acceptable risk to mean risk and cumulative harm on 2000 simulated scenarios.

of maximum acceptable risk used to guide AV trajectory and behavior planning. As a result, we showed how this leads to maneuver-based decisions, as motivated in the introduction. In contrast to other works that determine the AV behavior and evaluate safety measures, we specify a value for safety, namely the accepted risk, and observe the AV behavior as a result. Empirical evaluation reveals that maximum acceptable risk enhances safety even when risk is already minimized using a cost function. The best risk distribution is achieved using our two-stage cost function approach, which also shows the limits of risk management. Moreover, in this case, an empirical connection can be made between online risk in trajectory planning and a macroscopic view of risk, which could open new avenues into options for the safety argumentation of AVs. Considering a comprehensive approach to operational safety, which encompasses critical factors such as robust system reliability, well-designed fail-safe mechanisms, and seamless human-machine interaction, becomes imperative for successfully implementing the concept of maximum acceptable risk. However, this approach requires further steps, additional software uncertainties, and more data with multi-agent focus [37], to reveal the complex relationships in this area.

REFERENCES

[1] N. J. Goodall, "Machine ethics and automated vehicles," in *Road Vehicle Automation*. Cham, Switzerland: Springer, Jun. 2014, pp. 93–102. [Online]. Available: http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=P2DQHSYaCscUSLuo9CL&page=2&doc=11&cacheurlFromRightClick=no%0Ahttp://link.springer.com/10.1007/978-3-319-05990-7 http://link.springer.com/10.1007/978-3-319-05990-

[2] T. Fraichard. "Will the driver seat ever be empty?" 2014. [Online]. Available: http://hal.inria.fr/hal-00965176

[3] R. Benenson, T. Fraichard, and M. Parent, "Achievable safety of driverless ground vehicles," in *Proc. 10th Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, 2008, pp. 515–521.

[4] U.S. Dept. Transp. Office Secret. Transp., "Revised departmental guidance 2016: Treatment of the value of preventing fatalities and injuries in preparing economic analyses," vol. 4, no. 1, pp. 88–100, 2016. [Online]. Available: https://www.transportation.gov/sites/dot.gov/files/docs/2016 RevisedValueofaStatisticalLifeGuidance.pdf

[5] M. Geisslinger, F. Poszler, J. Betz, C. Lütge, and M. Lienkamp, "Autonomous driving ethics: From trolley problem to ethics of risk," *Philos. Technol.*, vol. 34, pp. 1033–1055, Apr. 2021. [Online]. Available: https://doi.org/10.1007/s13347-021-00449-4 https://link.springer.com/10.1007/s13347-021-00449-4

[6] "ISO 26262: Road vehicles-functional safety," Int. Org. Standard., Geneva, Switzerland, 2018.

[7] "ISO/PAS 21448: Road vehicles-safety of the intended functionality," Int. Org. Standard., Geneva, Switzerland, 2022.

[8] J. C. Gerdes, "The virtues of automated vehicle safety-mapping vehicle safety approaches to their underlying ethical frameworks," in *Proc. IEEE Intell. Vehicles Symp.*, 2020, pp. 107–113.

[9] S. Shalev-Shwartz, S. Shammah, and A. Shashua. "On a formal model of safe and scalable self-driving cars." 2017. [Online]. Available: http://arxiv.org/abs/1708.06374

[10] "Commission implementing regulation (EU) 2022/1426," Eur. Comm., Brussels, Belgium, Rep. 32022R1426, 2022.

[11] 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.

[12] C. Pek, S. Manzinger, M. Koschi, and M. Althoff, "Using Online verification to prevent autonomous vehicles from causing accidents," *Nat. Mach. Intell.*, vol. 2, no. 9, pp. 518–528, 2020. [Online]. Available: http://dx.doi.org/10.1038/s42256-020-0225-y

[13] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 903–918, Aug. 2014.

[14] M. Naumann et al., "On responsibility sensitive safety in car-following situations—A parameter analysis on german highways," in *Proc. IEEE Intell. Vehicles Symp.*, Jul. 2021, pp. 83–90.

[15] S. Vaskov et al., "Towards provably not-at-fault control of autonomous robots in arbitrary dynamic environments." 2019. [Online]. Available: http://arxiv.org/abs/1902.02851.

[16] J. Yoshida. "Robotaxi priorities: Avoid crashes or avoid blame?" 2022. [Online]. 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

[17] G. R. De Campos, R. Kianfar, and M. Brannstrom, "Precautionary safety for autonomous driving systems: Adapting driving policies to satisfy quantitative risk norms," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*. Indianapolis, IN, USA, Sep. 2021, pp. 645–652. [Online]. Available: https://ieeexplore.ieee.org/document/9564879/

[18] W. M. D. Chia, S. L. Keoh, C. Goh, and C. Johnson, "Risk assessment methodologies for autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 16923–16939, Oct. 2022.

[19] Y. Lin and M. Althoff, "CommonRoad-CriMe: A toolbox for criticality measures of autonomous vehicles," *IEEE Intell. Vehicles Symp. (IV)*, Anchorage, AK, USA, 2023, pp. 1–8, doi: 10.1109/IV55152.2023.10186673.

[20] A. Abdelhalim and M. Abbas, "A real-time safety-based optimal velocity model," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 165–175, 2022.

[21] M. Khonji, J. Dias, R. Alyassi, F. Almaskari, and L. Seneviratne, "A risk-aware architecture for autonomous vehicle operation under uncertainty," in *Proc. IEEE Int. Symp. Safe. Secur. Rescue Robot. (SSRR)*, 2020, pp. 311–317.

[22] T. Nyberg, C. Pek, L. Dal Col, C. Norén, and J. Tumova, "Risk-aware motion planning for autonomous vehicles with safety specifications," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2021, pp. 1016–1023.

[23] M. Emu, F. B. Kamal, S. Choudhury, and Q. A. Rahman, "Fatality prediction for motor vehicle collisions: Mining big data using deep learning and ensemble methods," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 199–209, 2022.

[24] K. Ren, H. Ahn, and M. Kamgarpour, "Chance-constrained trajectory planning with multimodal environmental uncertainty," *IEEE Control Syst. Lett.*, vol. 7, pp. 13–18, 2023.

[25] M.-Y. Yu, R. Vasudevan, and M. Johnson-Roberson, "Occlusion-aware risk assessment for autonomous driving in urban environments," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 2235–2241, Apr. 2019.

[26] *Ethics of Connected and Automated Vehicles: Recommendations on Road Safety, Privacy, Fairness, Explainability and Responsibility*. Paris, France: European Union, 2020.

[27] C. Luetge, "The german ethics code for automated and connected driving," *Philos. Technol.*, vol. 30, no. 4, pp. 547–558, 2017.

[28] L. Fraade-Blanar, M. Blumenthal, J. Anderson, and N. Kalra, "Measuring automated vehicle safety: Forging a framework," in *Measuring Automated Vehicle Safety Forging a Framework*. Santa Monica, CA, USA: RAND Corp., Oct. 2018. [Online]. Available: https://www.rand.org/pubs/research_reports/RR2662.html

[29] N. Kalra and S. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transp. Res. Part A Policy Pract.*, vol. 94, pp. 182–193, Dec. 2016.

[30] N. Kauffmann, F. Fahrenkrog, L. Drees, and F. Raisch, "Positive risk balance: A comprehensive framework to ensure vehicle safety," *Ethics Inf. Technol.*, vol. 24, no. 1, pp. 1–16, Mar. 2022. [Online]. Available: https://link.springer.com/article/10.1007/s10676--022-09625--2

[31] European medicines agendy. "PrOACT-URL." Accessed: Apr. 3, 2023. [Online]. Available: https://www.protectbenefitrisk.eu/PrOACT-URL.html

[32] S. Annell, A. Gratner, and L. Svensson, "Probabilistic collision estimation system for autonomous vehicles," in *Proc. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2016, pp. 473–478.

[33] Kraftfahrtbundesamt. "Verkehr in kilometern (VK)." 2021. [Online]. Available: https://www.kba.de/DE/Statistik/Kraftverkehr/VerkehrKilometer/vk_inlaenderfahrleistung/vk_inlaenderfahrleistung_node.html;jsessionid=F977BD0CEA15E1C2D70779E356A50678.live11312

[34] Statisitsches bundesamt. "Unfälle und verunglückte im straßenverkehr." 2021. [Online]. Available: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Tabellen/unfaelle-verunglueckte-.html

[35] D. M. DeJoy, "The optimism bias and traffic accident risk perception," *Accid. Anal. Prevent.*, vol. 21, no. 4, pp. 333–340, 1989.

[36] C. Starr, "Social benefit versus technological risk," *Science*, vol. 165, no. 3899, pp. 1232–1238, Sep. 1969. [Online]. Available: https://www.science.org/doi/10.1126/science.165.3899.1232

[37] P. Liu, R. Yang, and Z. Xu, "How safe is safe enough for self-driving vehicles?" *Risk Anal.*, vol. 39, no. 2, pp. 315–325, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/risa.13116

[38] Statisitsches Bundesamt. "Verletzte bei verkehrsunfällen nach art der verkehrsbeteiligung." 2021. [Online]. Available: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Tabellen/verletzte-fahrzeugart.html

[39] P. Dieppe, C. Bartlett, P. Davey, L. Doyal, and S. Ebrahim, "Balancing benefits and harms: The example of non-steroidal anti-inflammatory drugs," *Brit. Med. J.*, vol. 329, no. 7456, pp. 31–34, 2004.

[40] M. Werling, J. Ziegler, S. Kammel, and S. Thrun, "Optimal trajectory generation for dynamic street scenarios in a frenét frame," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 987–993.

[41] M. Geisslinger, P. Karle, J. Betz, and M. Lienkamp, "Watch-and-learn-net: Self-supervised Online learning for probabilistic vehicle trajectory prediction," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2021, pp. 869–875.

[42] M. Geisslinger, F. Poszler, and M. Lienkamp, "An ethical trajectory planning algorithm for autonomous vehicles," *Nat. Mach. Intell.*, vol. 5, pp. 137–144, Feb. 2023. [Online]. Available: https://doi.org/10.1038/s42256--022-00607-z

[43] "Crash report sampling system," U.S. Dept. Transp., Nat. Highway Traffic Safet. Admin., Washington, DC, USA, Rep., Accessed: Apr. 3, 2023. [Online]. Available: https://www.nhtsa.gov/crash-data-systems/crash-report-sampling-system

[44] "Automatisiertes und vernetztes fahren (Bundesministerium für verkehr und digitale infrastruktur)," Bundesministerium für Verkehr und digitale Infrastruktur, Ethik-Komission, Berlin, Germany, Rep., 2017. [Online]. Available: https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile

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

[46] M. Althoff, M. Koschi, and S. Manzinger, "CommonRoad: Composable benchmarks for motion planning on roads," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 719–726.

MAXIMILIAN GEISSLINGER received the B.Sc. and M.Sc. degrees in mechanical engineering and automotive technology from the Technical University of Munich, Munich, Germany, in 2016 and 2019, respectively, where he is currently pursuing the Ph.D. degree with the Autonomous Vehicle Lab, Institute of Automotive Technology. His research interests include ethics in autonomous driving, fair distribution of risk and quantification of uncertainties, and related applications for autonomous driving.

RAINER TRAUTH received the B.Sc. degree in engineering science and the M.Sc. degree in mechanical engineering from the Technical University of Munich in 2017 and 2020, respectively, where he is currently pursuing the Ph.D. degree in mechanical engineering with the Institute of Automotive Technology. His research interests include motion planning, situational awareness, and behavior planning approaches with a focus on real-world applications in autonomous driving.

GEMB KALJAVESI received the B.Sc. degree in mechanical engineering and the M.Sc. degree in production engineering from RWTH Aachen University in 2020 and 2022, respectively. He is currently pursuing the Ph.D. degree in mechanical engineering with the Institute of Automotive Technology, Technical University of Munich. His research interests include safety and self-awareness topics related to autonomous vehicles.

MARKUS LIENKAMP (Member, IEEE) received the degrees in mechanical engineering from TU Darmstadt and Cornell University and the Ph.D. degree from TU Darmstadt in 1995. He worked with Volkswagen as a part of an International Trainee Program and took part in a joint venture between Ford and Volkswagen, Portugal. He led the Brake Testing Department, VW Commercial-Vehicle-Development Section, Wolfsburg, Germany. He was the Head of the Electronics and Vehicle Research Department, Group Research Division, Volkswagen AG. He is a Professor with the Institute of Automotive Technology, Technical University of Munich, and is involved in the CREATE Project, Singapore. He is conducting research in the area of electromobility with the objective of developing new vehicle concepts. His research interests include advanced driver-assistance systems and vehicle concepts for electromobility. He has been the Chair of Automotive Technology at TUM since November 2009.