Safety and Risk—Why Their Definitions Matter

Nayel Fabian Salem, Sophie Le Page, Jason Millar, Philipp Junietz, Marcus Nolte, Robert Graubohm, and Markus Maurer

Contents

40.1 Introduction – 780

40.2 The Many Meanings and Uses of the Term 'Safety' – 781
40.2.1 How the Term 'Safety' is Used in Automotive Safety Engineering Frameworks – 782
40.2.2 How the Term 'Safety' is Used by Other Stakeholders – 791
40.2.3 Conclusions About the Meanings and Uses of the Term 'Safety' – 794

40.3 Questions About the Risks Associated with the Term 'Safety' – 795

40.4 Limitations and Recommendations – 796

40.5 Summary and Outlook – 797

References – 797

© The Author(s) 2026
H. Winner et al. (eds.), *Handbook Assisted and Automated Driving*,
https://doi.org/10.1007/978-3-658-45276-6_40

40.1 Introduction

The acceptance of automated road vehicles (amongst others) depends on their level of safety [1]. With regard to this generally accepted claim, it has been historically neglected that the discussion about the safety of automated vehicles requires further differentiation.

Surprisingly, compared to the research on automated driving in general, safety assurance of automated vehicles (cf. ► Chap. 46) is a relatively novel discipline. In 2004, the first DARPA Grand Challenge focused on automated driving and was held in the United States. Competing teams were challenged to design an automated automobile capable of navigating a 240 km course through the Mojave Desert. No team completed the course; the most successful entry completed a mere 11.78 km of the challenge. Even though no winner was declared, and even though automated driving technology has been a research topic in academia and industry since at least the 1980s [2–4], the 2004 Challenge is often cited as a landmark event in the evolution of automated driving technology. This is, in part, because a few years later in 2009 some of the participants joined Google to launch its self-driving car program. Google's project is significant because it captured the imagination of the public, sparking a public discourse around the goals for the safety of self-driving cars. Google's self-driving project was, and still is, hailed as an opportunity to significantly improve road safety by eliminating the human driver. Perhaps the most common statistic that one finds quoted alongside such claims is that over 90% of US road accidents are caused at least in part by driver error [5, 6]. Thus, since self-driving cars eliminate errors from the human driver, it is sometimes wrongly concluded that self-driving cars could eliminate up to 90% of road accidents. In light of such safety expectations it is not surprising that self-driving cars have received so much attention in the 13 years since Google launched its project. However, in the research community of automated vehicles the issue of safety assurance was underestimated for a long time. An example can be given by how safety was treated on the IEEE Intelligent Vehicles Conference in 2016 in Gothenburg. Some researchers suggested that safety was mostly an issue for series development of automated vehicles [7].

One important contribution to the acknowledgment of safety assurance for automated vehicles (especially in the German research policy) was made with the release of the book "Autonomous Driving: Technical, Legal and Social Aspects" [8] by giving recommendations for research demands in the field of automated vehicle safety. Publications from, for example, Schnieder and Schnieder [9], Reschka et al. [10], Bagschik et al. [11], Wachenfeld [12], and others [13–17] further highlighted the importance of safety in the field of automated driving. From a German perspective, these works set the stage for research on safety assurance of automated vehicles [18]. This trend was successfully followed by the initiation of publicly funded projects such as PEGASUS[1] in 2016. Subsequently, the PEGASUS project family was established and still contributes to the field of safety assurance research in Germany with projects such as "Verification and Validation Methods for automated vehicles Level 4 and 5"[2] (VVM) or "SET Level".[3]

In 2017 the German ethics commission "Automated and Connected Driving" elaborates that "the primary purpose of partly and fully automated transport systems is to improve safety for all road users" [19]. Consequently, the priority of safety in favor of utility was defined for all levels of automated driving as constraints of the technology. Whereas the statement of the ethics commission seems to be clear, implications for the development of automated driving systems are still unclear. Grunwald [1] pointed out that the trade-off between safety and utility is a key concern for the public acceptance of automated driving. For driver assistance systems (SAE Level 0–2 [20]) it was assumed that the driver finally takes the responsibility for the vehicle's behavior and hence also handles the trade-offs involved in participating in road traffic. However, automated driving systems (SAE Level 3+ [20]) will be required to perform this task on their own, which is a new challenge for the development and operation of such systems. The novelty of this challenge leads to ongoing questions. How safe do technical agents have to be that weigh more than one ton and move in public traffic with a speed of 30 m/s and more? Which mistakes are they allowed to make without losing general acceptance of automated driving? How does the paradigm of "vision zero" and the sustained claim of research and industrial institutions of the feasibility of zero fatalities and severe accidents in road traffic influence public expectations towards automated driving? Which significance for the acceptance of automated vehicles does the common language understanding of safety as the "freedom from harm or danger" (Merriam-Webster Dictionary) [21] have, as it implies the complete absence of risks that are associated with "safe" technical systems? Please note, that for the sake of simplified public communication we do not distinguish between automated driving systems and automated vehicles in the context of this chapter. We acknowledge that the definition of the system boundary influences aspects that should be considered for system safety. However, for the fundamental discussion of automated driving safety we intentionally neglect this issue.

1 ► https://www.pegasusprojekt.de/en/home.
2 ► https://www.vvm-projekt.de/en/.
3 ► https://setlevel.de/en.

In contrast to the definition given in the Merriam-Webster Dictionary, safety engineering standards such as ISO 26262:2018 and IEC 61508:2010 define safety as the "absence of unreasonable risk" or "freedom from unacceptable risk". The semantic difference between the common definition of 'safety' and those definitions provided by engineering standards depend on what is considered "unreasonable". In safety engineering standards a clear definition of unreasonable risk is not provided. However, the guidelines in these standards in terms of arguing for safety always rely on the definition of unreasonable risk. Based on this vagueness it becomes clear that the meaning and use of the term 'safety' is of fundamental importance for assuring and hence arguing for the safety of an automated driving system (SAE Level 4+). The public acceptance of a provided argumentation of safety will finally have a major influence on the acceptance of the technology [1]. It is already acknowledged in the standardization of "safety of the intended functionality" of automated road vehicles [22] that meeting a risk acceptance criterion is required in order to deploy these systems. However, the definition of such publicly accepted criteria is still an open challenge [23] (cf. ▶ Chap. 6). Additionally, even the metrics for measuring safety are not established yet, which however is a necessary step before defining acceptance criteria.

One could ask why the fundamental discussion of safety and risk is relevant now and why it was successfully neglected (rightly or wrongly) for driver assistance systems. Unfortunately, current road traffic involves a number of fatalities and severe injuries. However, this fact seems to be accepted by the public as an inherent risk of road traffic that can be reduced but not completely eliminated [24, 25]. Such an inherent risk being accepted is far from obvious. It is the result of ongoing negotiations between the objectives of road safety and mobility of involved stakeholders. For conventional road vehicles in Germany there has been a clearly decreasing amount of accidents over the past decades [14]. However, it is unclear how this evaluation will be conducted for automated vehicles (SAE Level 3+) as a part of road traffic [17].

The aim of this chapter is to point out the lack of a common understanding of the term 'safety' where automated vehicles are concerned and to give examples of the different meanings and uses of the term from the perspectives of automotive development engineers, regulators, legislators, advocacy groups, and the public. Please note that this analysis shall not represent any of the included stakeholder groups exhaustively. Thus, we seek to provide a description of the problem based on the selected examples rather than attempting to propose solutions that harmonize the term 'safety'. Furthermore, we will provide guidance for a more cohesive discussion of safety across disciplines and stakeholder groups. One could ask why differences in the meanings and uses of 'safety' matter after all, since the more pressing question, at least in the automotive engineering community, has been "How safe is safe enough?" for some time now. We argue that the inability to provide an answer to this question by merely focusing on the engineering aspects of the problem is also based on the different answers to the question across all relevant stakeholders of: "What does safety (of automated vehicles) even mean?" Hence, the focus of this chapter is to provide evidence for the plethora of possible answers to the latter question.

Discussing safety engineering in the context of automated driving and pointing out its associated risks is often (mis)interpreted as an impeding activity. On the contrary, in this chapter we contribute to the goal of enabling sustainable acceptance of automated vehicles as participants of road traffic. Hence, we focus on showing differences in the meaning and use of the term safety between different stakeholders in ▶ Sect. 40.2. Based on our analysis we pose questions that could help facilitate an explicit and more detailed distinction between aspects of risk considered by different stakeholders in ▶ Sect. 40.3.

40.2 The Many Meanings and Uses of the Term 'Safety'

In the current landscape of standards for the safety assurance of automated driving systems (cf. ▶ Chap. 46) the term 'safety' is more specifically considered by focusing on aspects of functional safety and safety of the intended functionality (SOTIF) (cf. ▶ Chap. 6). Definitions for these specific aspects of safety are provided by ISO 26262:2018 (functional safety) and ISO/DIS 21448:2022 (SOTIF). Functional safety thereby is defined as the "absence of unreasonable risk due to hazards caused by malfunctioning behavior of E/E systems" [26] whereas SOTIF is referred to as the "absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or its implementation" [22].

Taking a more general perspective, the term 'safety' tends to refer to a state in which there is an absence of unreasonable risk [22–29]. Applying that generalized meaning of safety to automated vehicles (SAE Level 3+), a safe vehicle would be a vehicle that poses no unreasonable risk. Notice, however, the degree of ambiguity and vagueness included in that general sense of the term: it is unclear who, precisely, is otherwise at risk. It could be drivers, passengers, pedestrians who are exposed to risk. It is unclear what amount of risk would be considered unreasonable for any particular hazard. It is further unclear what specific hazards are bundled in the concept of safety. Are the hazards of physical na-

ture? Are they associated with psychic or even environmental aspects? These details matter since to design a "safe" vehicle, or to designate a vehicle as "safe", one must agree on what, specifically, is meant by "safe".

It is important to note from the outset that safety critical functionality as a term encompasses specific risks. Determining what is understood as "safe" thus requires to clarify the kind of risks being discussed, what would be a reasonable, or unreasonable amount of risk, and who ends up being safe once all that is sorted out. This is no easy task as we will discuss in this section. There are many ways that the terms "safe/safety" are currently deployed in the context of automated vehicles. Currently, there is no consensus among stakeholders, be they engineers, regulators, legislators, advocacy groups, or the public, what would constitute a safe automated vehicle. As such, different stakeholders could mean fundamentally different things when they raise concerns or make claims about the safety of automated vehicles. This is a problem for all these stakeholders because their diverging use of the concept of "safety" could set up uneven expectations of how automated vehicles will behave. Engineers using separate definitions of safety to design vehicles could result in vehicles that behave very differently from one another, and differently from how other stakeholders, the public included, expect them to behave. Ultimately, mismatched expectations of this sort could impact the actual or perceived trustworthiness of automated vehicles, possibly leading to lower trust among regulators, legislators, advocacy groups, and the public and thus could lead to problems related to acceptance.

In this section, we analyze how different stakeholders understand and conceptualize safety by breaking down safety statements into their associated risks. Since engineering frameworks (in industry or research) for designing safe automated vehicles are explicit about associated risks in their definitions of safety, we summarize their use of the term 'safety' into the following safety statements: "A safe automated vehicle poses no [insert risk]". In contrast, other stakeholder groups such as legislators, regulators, advocacy groups, and the public media are less explicit about the risks associated with their safety statements about AVs. Therefore, we found that it would not improve the clarity of this section to include similar safety statements for other stakeholder perspectives. Please note that these chosen stakeholder perspectives are not necessarily distinctive but enable us to highlight their differences. Hence, our summaries of safety statement shall not represent the perspective of a stakeholder group exhaustively but give an example of how the term 'safety' is used sometimes.

In order to elaborate the many meanings and uses of the term 'safety' we first analyze several automotive safety engineering frameworks provided by standards or proposed in the research community (▶ Sect. 40.2.1). With our analysis we provide examples of how the concept of safety is used among automotive engineers as one relevant stakeholder group directly influencing the design of automated vehicles. We then survey several sources to gain some understanding of how "safety" is used among other stakeholder groups (▶ Sect. 40.2.2), which directly or indirectly influence, or are influenced by, the design of automated vehicles such as regulators, legislators, corporations, advocacy groups (e.g. Mothers Against Drunk Driving (MADD)), and public media. What we find in our analysis is that safety and its associated risks are talked about differently between stakeholder groups, which may lead to mismatched expectations of what would constitute a safe vehicle. A synthesis of all possibly considered risks is not in the scope of this chapter as it would require a more thorough review of existing literature as well as qualitative and quantitative studies on relevant stakeholder perspectives. With the identification of a diverging meaning and use of the term 'safety' we aim for contributing to a more explicit discourse of risks without assuming consensus on what constitutes safe automated driving.

40.2.1 How the Term 'Safety' is Used in Automotive Safety Engineering Frameworks

In this chapter, we provide indications to support the hypothesis of mismatched meanings and uses of the term 'safety' for automated vehicles with regard to the risks that are found to be relevant for different stakeholder groups. In this section, we focus on different understandings of safety in the automotive industry and engineering research community. First, we give an example of how safety is interpreted in normative automotive safety engineering frameworks (▶ Sect. 40.2.1.1). Additionally, we show how suggested safety engineering frameworks that were published by corporate and academic research groups use the term "safety" differently. As two examples of such suggested frameworks, we elaborate how safe automated vehicles could be understood by analyzing formal safety verification approaches (▶ Sect. 40.2.1.2) and how one could include human values as a part of the definition of safety following an approach of value-sensitive development (▶ Sect. 40.2.1.3).

40.2.1.1 Established Automotive Safety Engineering Frameworks

Functional Safety and SOTIF

For the development of driver assistance systems, standards for functional safety (e.g. ISO 26262:2018)

and SOTIF (e.g. ISO 21448:2022) provide guidelines to assure these different aspects of safety. In order to analyze the meaning and use of the term 'safety' in both of these safety engineering frameworks, we provide an example in the scope of functional safety. Since the terminology used for SOTIF in the ISO/DIS 21448:2022 often times refers back to definitions from functional safety, we focus on analyzing terminology provided in the ISO 26262:2018.

The ISO 26262:2018 is an international series of standards that provides a reference framework for the automotive functional safety lifecycle. In the standard, there is a very particular understanding of safety and risk. According to the ISO 26262:2018 risk is a "combination of the probability of harm and the severity of that harm" (which is compatible with Bayesian risk [30]). Harm is referred to as "physical injury or damage of the health of persons". In order to quantify physical injuries, the ISO 26262:2018 suggests using for example the abbreviated injury scale (AIS). However, the exact extent of harm is hard to determine [31]. Hence, severity is defined as an "estimate of the extent of harm to one or more individuals that can occur in a potentially hazardous event". A hazardous event is considered as the "combination of a hazard and an operational situation". In order to separate any possible harm from the harm that is considered within the scope of the ISO 26262:2018, a hazard is defined as a "potential source of harm caused by malfunctioning behaviour of the item". The item would in this case be an electric or electronic subsystem, which is potentially part of a larger system. Furthermore, the probability of harm was mentioned in the definition of risk. This probability is constituted by the exposure to the operational situation and the probability of a system failure. The exposure of a hazardous event refers to the "state of being in an operational situation", which again is defined as a "scenario that can occur during a vehicle's life". With these terms being introduced in the ISO 26262:2018-1 [26], there is a clear focus on physical injury as the kind of risk being addressed by the standard.

In order to develop functionally safe systems based on this terminology, the ISO 26262:2018 demands to determine automotive safety integrity levels (ASIL) (cf. ▶ Chap. 6) based on a classification of identified hazards and a subsequent risk assessment. The parameters of the risk assessment are the probability of a "state of being in an operational situation" (exposure), the "estimate of the extent of harm to one or more individuals that can occur in a potentially hazardous event" (severity), and the potential for the "ability to avoid a specified harm […] through timely reactions of the persons involved" (controllability). After identifying a hazard, it can be assessed by using these three parameters. The system design subsequently has to fulfill the derived goals with an ASIL that is based on this classification.

As an illustration of a hazard analysis and risk assessment in the concept phase of the safety lifecycle and its implications for the underlying meaning of safety, we provide an example of a safety analysis of an automated driving system that shall be operated in SAE Level 4. As the focus of this example is not the exact result of the safety analysis, but rather the workflow, we generously simplify multiple process steps.

For the operational design domain, we assume rural roads with a maximum allowed speed of 100 km/h. One hazard could be that the vehicle leaves the road and crashes into one of the neighboring trees. A possible reason could be a malfunction that leads to the complete loss of the steering function. An operational situation of driving along a curved road sequence with trees on both sides of the road has a high probability (exposure class E4) as it is highly probable (more than 10% of average operation time in the selected target market [31]) for rural roads to have neighboring trees. For the severity of the hazard, we estimate at least life threatening injuries (severity class S3) since even after executing an emergency braking maneuver, the impact speed for a frontal collision with a neighboring tree would be more than 40 km/h (suggested minimum speed for S3 by SAE J2980 [32]). As there are no persons involved in the operational situation (the system is operating on SAE Level 4) that could contribute to the controllability of the hazardous event we have to assess it as uncontrollable (controllability class C3). Based on this example of a hazard analysis, we take the recommendation from the ISO 26262:2018 and derive a required ASIL D for the preservation of the steering function. In the ISO 26262:2018 it is suggested that for example for an ASIL D function, a random hardware failure target value of less than $10^{-8}\,h^{-1}$ has to be met. This target value would be applied to hardware components in later steps of the safety lifecycle. However, with respect to the definition of safety as the absence of unreasonable risk this step is not straightforward. What is implicitly involved in determining an ASIL is first of all a risk assessment of the classified hazard. Afterwards, this estimated risk is compared with an acceptance criterion. If the risk involved in the analyzed hazard is higher than required by the risk acceptance criterion, measures to reduce the risk of the hazard are taken. These measures are specified using safety integrity levels and are applied to software and hardware components in subsequent phases of the safety lifecycle.

Based on our example, we argue that the underlying meaning of safety in the ISO 26262:2018 for this specific hazard and functionality is that the system is safe, if among others the design goal of a failure rate of (in this example) less than $10^{-8}\,h^{-1}$ is met, because it is defined as reasonable by a risk acceptance criterion. However, why is the design goal not $10^{-12}\,h^{-1}$? Since functional safety has a legacy of systems to refer to, its risk

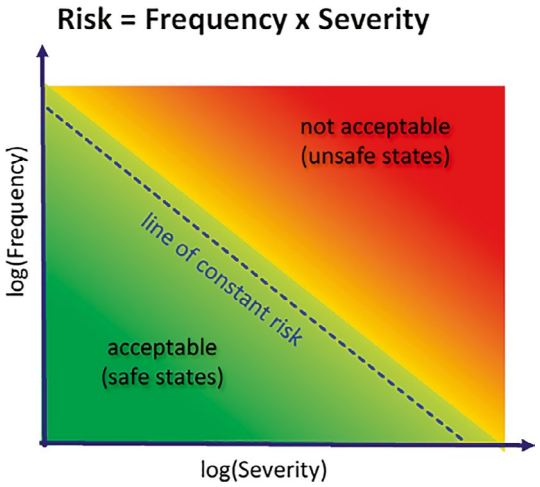

Fig. 40.1 Illustration of risk

acceptance criterion is based on a mix of proven-in-use observations and assumptions based on accident research. For the assurance of automated vehicle safety especially considering SOTIF on the other hand, it is not clear yet, which risk acceptance criterion should be chosen.

In the following section (▶ Sect. 40.2.1.1.2), we present in more detail how quantitative risk assessment is implicitly performed in the established safety engineering frameworks, explain differences in possible risk acceptance criteria such as ALARP, MEM and GAMAB, and elaborate what considerations could be included in choosing one over the other acceptance criterion.

Quantitative Risk Assessment[4]

The usual quantitative risk definition "risk equals frequency times severity" is illustrated in Fig. 40.1. Due to the double logarithmic scale, a line represents constant risk.

Considering the occurrence of unintended events (frequency) and extent of damage (severity), ideally two areas are defined. In the green area, the system is in a safe state; the corresponding risk is accepted. In the red area, the system is in an unsafe state; the corresponding risk is not accepted. The borderline between these two areas is probably not sharp; there can be a kind of transition area in between.

Although this simple definition is very useful for many questions in technology and insurance industry, it neglects aspects like aversion against high severity, lack of controllability, and personal benefit, which are relevant for risk perception and acceptance by individuals and society. Intensive research on risk perception and risk acceptance started in the second half of the last century. Different authors analyzed risk acceptance and risk–benefit constellations in various studies (Douglas [33], Crouch [34], Gibson [35] Grunwald [1], Kinchin [36], Kletz [37], Starr [38, 39], and Webb [40]). Fritzsche discussed risk acceptance relating to voluntary nature of exposure based on the above-mentioned studies [41]. Slovic concludes similar results in a more recent, updated publication [42]. It is interesting that both authors conclude similar risk numbers despite the major gap of several decades. The reason might be that risk perception studies reached their peak in the 1970s with the introduction of nuclear power. Junietz et al. [43] suggest correcting the numbers with a factor derived from the change in mortality rate.

For voluntary activities, Fritzsche [41] found that the willingness to accept risks is nearly unlimited, depending on the experienced personal benefit. We can see this by the example of high-risk sport or other leisure activities, e.g. free climbing or motorcycling. Job-related activities are important for a deeper understanding of the subject. Acceptance is relatively well investigated in this field and there is a common understanding of accepted individual mortality risk in the order of 10^{-5} per person and year, for example by professional associations and insurance companies, on the one hand. On the other hand, job-related risks are useful to bridge the gap between voluntary and involuntary risks.

Fritzsche [41] found that for involuntary risks, e.g. death of passengers due to a train or airplane crash, the acceptance level is an order of magnitude lower than for job-related risks. Moreover, acceptance decreases another order of magnitude, if the risk is caused by major technology, e.g. chemical industry or nuclear power generation. Beside the fact that the experienced personal benefit of those technologies is low (at least from a subjective point of view), the low degree of self-determination or rather controllability by individuals plays an important role for the low acceptance level as well as the potentially high number of mortalities (severity). Nevertheless, the summarized studies show that it is generally possible to deal with risk, risk perception, and acceptance in a quantitative manner.

To implement safety requirements based on risk acceptance, several concepts have been developed in different application areas. As there is a relationship between railway and road traffic, it is useful to refer to the CENELEC safety standard EN 50126. The development of this standard has been started in the 1990s where safety requirements based on quantitative risk analysis have been implemented and ALARP, MEM, and GAMAB have been introduced as principles for risk acceptance. Those principles are briefly explained in the following sections.

4 This section is largely adopted from Junietz [17].

Table 40.1 Accidents on German highways [47]

Severity	ISO 26262 severity level	Average distance between two accidents of this level (km)	Accident rate per driven distance
Fatal	S3	660×10^6	1.52×10^{-9}
Severe injuries	S2	53.2×10^6	1.88×10^{-8}
Injuries	S1	12.5×10^6	8.00×10^{-8}
W/O injuries	S0	7.5×10^6	1.33×10^{-7}

As Low as Reasonably Practicable (ALARP)

ALARP tries to assess what is technically feasible considering economic sense and socially acceptable. Between the two regions of generally unaccepted and broadly accepted risk, there is a tolerance range where risk is undertaken only if a benefit is desired and where each risk must be made as low as reasonably practicable.

Derivation of risk figures is not directly applicable as EN 50126 failed to give certain values for generally unaccepted and broadly accepted risk. However, other authors, for example Risk & Reliability Associates, deliver both values [44]: The two key levels seem to be located around road death statistics (about 10^{-4} per person and year) and the chances of being struck by lightning (about 10^{-7} per person and year). If something is more dangerous than driving a car, the risk is unacceptable. If something is less dangerous than being struck by lightning, it is not required to be reduced further. In the range between these two figures, cost benefit studies are appropriate to reduce the risk to as low as reasonably practicable. Especially this lower ALARP limit corresponds very well with the acceptance criterion for major technology risks shown in Fig. 40.3.

Minimum Endogenous Mortality (MEM)

MEM is based upon age- and gender-specific mortality rates [45]. Although the absolute values of the mortality rates change with birth cohort, they show a typical development over age as well as a significant minimum at an age of about 10 years, which is about $5 \cdot 10^{-5}$ per person and year. The related mortality at an age of 10 years is defined as "minimum endogenous mortality".[5] The MEM principle demands that a new system does not significantly contribute to the existing minimum endogenous mortality. EN 50126 specifies that the individual risk due to a certain technical system must not exceed 1/20th of the minimum endogenous mortality, considering that people are normally exposed to the risk of several technical systems. This means that the acceptable individual risk of a certain technical system is $2.5\ 10^{-6}$ per person and year, when using latest mortality rates as a basis (EN 50126 uses mortality rates from the 1980s). This value corresponds very well with the acceptance criterion for involuntary risks shown in Fig. 40.3.

Globalement Au Moins Aussi Bon (GAMAB)[6]

GAMAB, (or GAME globalement au moins équivalent), requires, unlike MEM, the existence of a reference system with—currently—accepted residual risks. According to GAMAB, residual risks caused by a new system must not exceed those of the reference system. That is, a new system must offer a level of risk generally at least as good as the one offered by any equivalent existing system. Hence, it is necessary to identify the risk of an equivalent existing system.

Looking for the acceptable risk of automated driving (SAE Level 3+) in a certain application area according to GAMAB, the current risk of an equivalent existing system in the same application area needs to be identified. To derive acceptance requirements for a controlled-access highway pilot, the current risk on German highways (so called "Autobahn") during manual driving is analyzed. Table 40.1 shows average distances between two accidents referring to severity levels according to ISO 26262 [46]. However, for other Level 3+ systems, there might be no reference system as today's vehicle fleet does not cover the use-case.

Figure 40.2 shows the observed accident rates on German highways versus severity levels according to the ISO 26262. Comparing risks of the different severity categories requires weighting of the different levels. However, there is no standardized way (see also Baum [48], Hydén [49], and Wachenfeld [12]). The lines of constant risk in Fig. 40.2 assume that the difference between adjacent severity levels is one order of magnitude, or in other words that an accident with fatalities is assessed with a ten times higher severity than an accident with severe injuries. However, that does not mean that a fatal accident is definitely ten times worse. The factor of ten fits to the current accident frequency, which could change in the future e.g. due to enhanced passive safety. However, this assumption allows defin-

5 Although the mortality statistics does not differentiate between endogenous and exogenous causes for the death.

6 English: generally at least as good as.

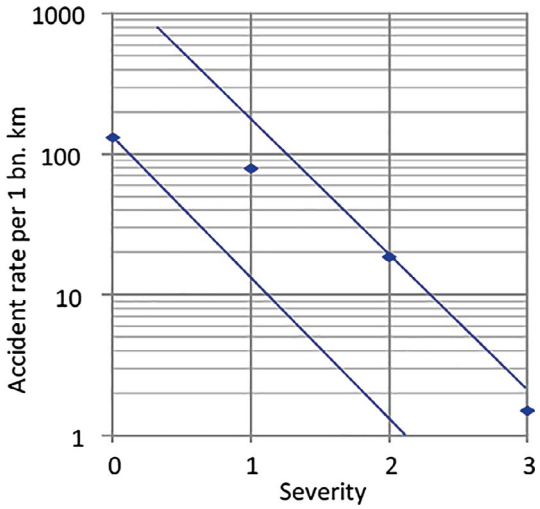

Fig. 40.2 Quantitative accident risk on German highways [47, 50]

ing a band of constant risk in current traffic, which can be used as reference. As already discussed, the risk will not be accepted above the upper envelope. Beyond the lower envelope, the risk might be accepted. Between both lines, there is a transition area. In accordance with the ALARP principle, this is a tolerance zone where risk is undertaken if a benefit is desired and where each risk must be made as low as reasonably practicable.

In Fig. 40.3, the results of the application of the different risk acceptance principles are displayed related to the risk acceptance limits of the different expositions explained previously. The different approaches for the mortality risk are summarized. On the one hand, it shows that the application of different risk acceptance principles delivers comparable and consistent results. On the other hand, it demonstrates that we have to deal with a relatively broad range of applicable acceptance criteria.

Considering the impact of voluntary exposure, different groups of users have to be distinguished, for example users of automated driving systems and other traffic participants. Finally, a comparison with other technologies—especially other traffic systems and technologies that deliver a high personal benefit—seems to be useful. Additionally, a decrease in total mortality risk is expected in the future, following the trend in the last decades and centuries. Therefore, risk acceptance might change over time.

Summary

In this part of our analysis, we gave examples of normative frameworks in automotive safety engineering and examined how they describe and address safety. There seems to be an agreement in frameworks like functional safety (ISO 26262:2018) and SOTIF (ISO/DIS 21448:2022) that the risks associated with a safe system are related to physical injuries. As we will do for the other frameworks and perspectives we consider in this chapter, we summarize the respective meaning of safety for automated vehicles:

> A safe automated vehicle poses no unreasonable physical harm to persons involved in an operational situation; unreasonable is defined as an agreed upon maximum probability of specific severities of harm.

Additionally, we pointed out that safety engineering involves including retrospective acceptance criteria like for example a number of acceptable injuries. Such criteria are used as part of the assessment of prospective estimations in terms of exposure, severity and controllability. Within this process, there is also the task of changing the scope in the analysis. Risk acceptance criteria refer to the entire operational environment on a macroscopic level. For example, the

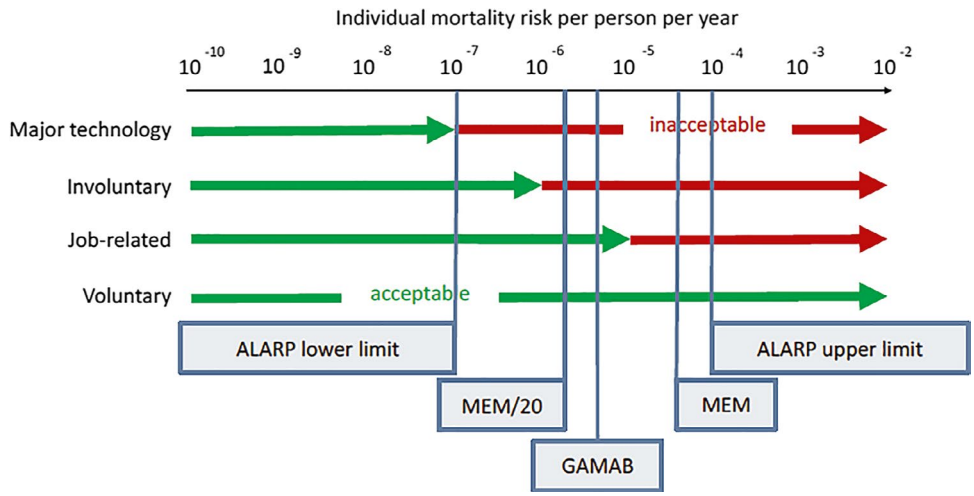

Fig. 40.3 Application of different risk acceptance principles to highway accidents, translated from Steininger and Wech [51], based on Fritzsche [41], GAMAB is based on the risk on German controlled-access highways and MEM is based on the current mortality tables

German Ethics Commission on automated and connected driving raised the goal of a positive risk balance [19] and generally refers to the guiding principle of avoiding accidents. However, it is still unclear how to account for this goal in every considered scenario [52] on a microscopic level. What the positive risk balance already defines is the reference for what makes an automated vehicle safe. A positive risk balance or "safety" is thus achieved when an automated vehicle introduces less risk with respect to accidents than human drivers.

40.2.1.2 Frameworks for Formal Safety Assurance

As automated driving introduces new challenges for safety assurance in an open traffic context (cf. ▶ Chap. 46) several possible solutions have been proposed in the research community. We present one group of approaches for assuring safety in an open traffic context in this section. The concept of formally proving safety is generally realized by agreeing on basic rules and implement their formal representations as part of the safety assurance process. Approaches that realize such a safety assurance strategy were presented for example by Shalev-Shwartz et al. [53] with Responsibility Sensitive Safety (RSS), Nistér et al. [54] with Safety Force Fields (SFF) or Althoff [55] with the utilization of reachable set analysis. For an in-depth presentation of the approaches, we refer the reader to the respective cited sources. In this section, we focus on the underlying meaning of safety and thus only provide an overview of the selected approaches. As an illustration, we will furthermore provide an example of how the interpretation of formally provable safety is manifested in a scenario using the RSS framework.

In their introduction of Safety Force Fields (SFF) Nistér et al. [54] limit their notion of safety to the avoidance of colliding with obstacles. By conceptually separating the nominal behavior planner from a distinct safety behavior planner that could rely on Safety Force Fields, they argue that a vehicle would never contribute to an unsafe situation. Given this condition, Nistér et al. [54] define: "If all actors do what they are required, no collisions can occur." Hence, this use of the term 'safety' is clearly limited by the acceptance of the assumption, that a safe vehicle is one that never contributes to an unsafe situation.

Althoff [55] relies on a similar notion of safety: "A major constraint for the trajectory planner is that the generated trajectories have to be safe, i.e. static and dynamic obstacles must not be hit." However, in his approach Althoff [55] proposes to use stochastic models (reachable sets) to proof that no collision is possible in a given situation. When comparing the notion of safety from Althoff's work [55] to Nistér et al. [54], it becomes clear that both offer closed solutions for handling safety in an open context by relying on necessary assumptions for their proven models. Making assumptions in order to formally prove safety (or any other property for that matter) cannot be attributed to their specific approaches but is an inherent challenge of using closed mathematical models in an open world.

Shalev-Shwartz et al. [53] introduce Responsibility Sensitive Safety (RSS) and acknowledge necessary trade-offs for automated vehicles acting in an open traffic context. In comparison to established safety engineering frameworks Shalev-Shwartz et al. [53] propose that instead of identifying particular hazards and quantifying the associated physical risks, five "common sense" rules can be formulated and proven mathematically. There are multiple implications of this strategy with respect to the meaning of safety. First, there is an implicit representation of the risks that are considered by RSS. High-level rules try to generalize different kinds of risks to different stakeholders. Therefore, the traceability of stakeholder concerns and development assumptions to their implementation in the behavior of the automated vehicle in road traffic is limited. Second, RSS requires stakeholders to agree on the parameters (just like the assumptions made by Nistér et al. [54] and Althoff [55]) that are used in the formal proof of such kind of safety. Finally, when automated vehicles would be deployed adhering to RSS, then the argument by Shalev-Shwartz et al. [53] suggests that in case of an accident where an automated vehicle behaves compliant with the defined RSS parameters, the responsibility would not be ascribed to the vehicle or respectively the car manufacturer. In contrast to established automotive safety engineering frameworks, there is no explicit integration of risk acceptance criteria. We show what this means for the understanding of safety using RSS by giving a simplified example.

In ◘ Fig. 40.4 Error! Reference source not found. the blue automated vehicle drives from left to right and passes a row of parked vehicles. Oncoming traffic is represented by the red vehicle and added to limit the degrees of freedom of the blue vehicle to the longitudinal speed. Designing a vehicle for such scenarios includes trading of mobility and risk [25]. Merely obeying the speed limit could mean that the mobility of the passengers is as high as the legal traffic code allows. However, by including the uncertainty about a child potentially running on the street between the parked cars and onto the road makes it difficult to argue merely for legal responsibility. In Germany, the argumentation would be even more challenging as the "principle of legitimate expectations" [translated by Salem, Vertrauensgrundsatz] does not apply to children and persons, who are obviously impaired with respect to acting in or perceiving their environment. The uncertainty in the open traffic context leads to an inherent risk of

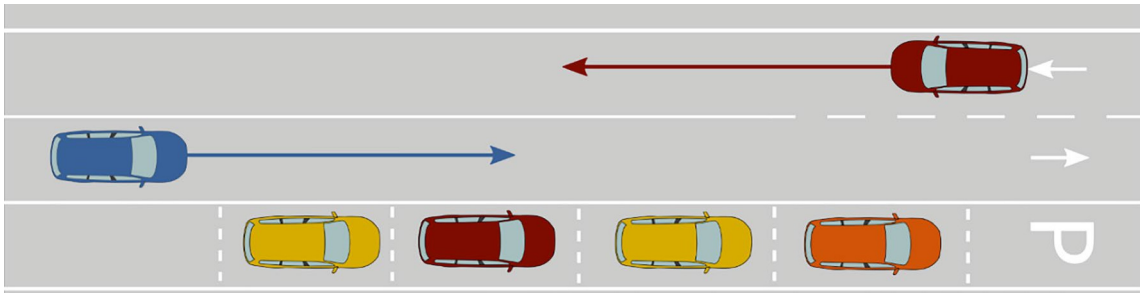

Fig. 40.4 Example of a scenario where the automated vehicle passes a row parked vehicles with possibly occluded pedestrians

participating in road traffic. In the scenario, this uncertainty is represented by the potentially crossing child behind one of the parked vehicles.

RSS proposes a framework for engineers and regulators to agree on a parametrization (in this scenario implying a certain speed) that is associated with an accepted residual risk. RSS acknowledges that trade-offs have to be made due to uncertainties in the open context. However, a consequence of applying the RSS framework to this scenario implies that there is a speed that would exempt the blue vehicle from any (legal) responsibility for an unavoidable collision with a crossing child. Hence, RSS assumes that the residual risk in traffic resulting from the chosen parameters is accepted by all stakeholders.

Even though each of the selected approaches towards formal safety assurance suggest different frameworks to prove system safety, they generally rely on a similar meaning of safety. Implicit trade-offs with respect to the risks considered are a fundamental assumption that approaches such as RSS, safety force fields or reachability analysis make by their very nature of attempting a formal proof of safety. Based on our analysis, we suggest an explicit summary of how the term 'safety' is used in these frameworks:

> A safe automated vehicle is one that poses no risk of violating a formally proven model; the validity of the model is based on agreed assumptions.

This poses challenges for drawing conclusions for our initial questions. Who is at risk? What kinds of risks are addressed? With respect to which reference is the system safe? However, with regard to mismatched expectations between stakeholders it seems important to be explicit about assumptions and trade-offs that are made with respect to safety [23].

40.2.1.3 Value-Sensitive Development Framework

While formal methods for proving safety considered in the context of this chapter do not offer explicit trade-offs, in this section we introduce a framework for value-sensitive development that acknowledges the importance of making trade-offs with respect to safety and risk in an explicit manner. From the perspective of value-sensitive development, explicitly weighing values is important because the risks that are associated with the meaning of safety otherwise are unclear to the relevant stakeholders. Hence, value-sensitive development includes stakeholder values explicitly and provides methods to design for safety, while acknowledging risks of violating stakeholder values when making trade-offs. As the term 'value' is associated with terminological discussions, we do not attempt to define the term universally but we rather show how the term is used within a value-sensitive development framework.

A common definition of the term 'value', which is also referred to by Ulbrich [56] in his work on value-sensitive behavior planning, is given by Kluckhohn [57].

> » A value is a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable which influences the selection from available modes, means, and ends of action. [57]

What remains unclear is, whether the 'conception' of a value has an absolute nature or if it can also be expressed relatively as a result of a weighing process. Albert [58] elaborates the differences concerning the terminology of a "value" from a psychological, sociological and cultural science perspective and proposes a more detailed definition.

> » A value as a measure with an adaptable weight and arbitrarily in relation with further measures alike constitutes a value pattern that reduces complexity for stakeholders and determines actions within situational possibilities variously direct. (translated by Salem, based on Albert [58])

In the context of value-sensitive development this definition of values accounts for the need of weighing stakeholder values depending on the situational context as it differentiates between value patterns and values. Hence, based on the definition by Kluckhohn [57] and Albert [58], values can be understood as objectives

desired by stakeholders that can be weighed against each other in a given context.

After having discussed the interpretation of the term 'value' in the context of value-sensitive development, we will focus on its relation to the term 'safety' and 'risk'. Value-sensitive development, just like established automotive safety engineering frameworks (▶ Sect. 40.2.1.1), interprets safety as the absence of unreasonable risk. However, the argumentation for reasonable design decisions is not enabled by fulfilling an abstract risk acceptance criterion but by explicitly documenting trade-offs between stakeholder values and formally representing them throughout the product lifecycle of an automated vehicle. In contrast to for example the work by Thornton [59] and Ulbrich [56] safety is therefore not understood as a value but as a result of weighing the violations of stakeholder values.

> A safe automated vehicle is one that poses no unreasonable risk of violating stakeholder objectives (or values).

As a result of associating safety with a reasonable risk of violating stakeholder values, risks can not only be related to physical harm (cf. ▶ Sect. 40.2.1.1) as stakeholders could desire more than physical wellbeing. This notion of safety including risks beyond physical harms is also in line with the kinds of risks included in the concept of safety in the UL 4600, and ISO 31000:2018. Risk as "a combination of the probability to occurrence of a loss event and the severity of that loss event" is defined in the UL 4600. A loss is further defined as "human death, human injury, animal death, animal injury, property damage, environmental damage, or other substantive adverse outcome". With regard to the losses that should be considered when designing safe automated vehicles, UL 4600 enables the discussion of what could be understood as "substantive adverse outcome". The example of including financial cost caused by a failure already indicates that risks beyond physical nature could be included in the safety case.

ISO 31000 further defines risk as the "effect of uncertainty on objectives". This definition enables an even broader consideration of risks that should be addressed in a safety case for automated vehicle. However, note that ISO 31000 is originally intended as a guideline for risk management in organizations, where the assessment of, for example, fatal injuries plays a subordinate role. We apply this meaning of safety to the context of automated vehicles in order to show a potential mismatch within organization's risk management as they introduce this technology.

As a result, in value-sensitive development the meaning of the term safety is associated with risks related to the objectives of stakeholders such as mobility, physical wellbeing, or legal compliance. Because these values or objectives are represented explicitly in the design process it is possible to trace them to the behavior of an automated vehicle. Explicitness of values enables stakeholders to evaluate, whether the behavior could be considered safe from their perspective. However, stakeholder values are highly context dependent. That is, whereas in the scope of one scenario a value trade-off could be considered safe, it might be found unsafe within the scope of the entire operating environment when considering (possibly additional) stakeholders and their values. Hence, a clear definition of the scope is necessary when making trade-offs between values.

In order to describe challenges of integrating the notions of values and safety and since value-sensitive development is not an established framework for handling safety, we introduce essential activities that are part of a value-sensitive development framework (◘ Fig. 40.5) and afterwards reflect the respective meaning and use of the term 'safety'.

The foundation of value-sensitive development is the acknowledgement of stakeholder values that shall be represented in the behavior of an automated driving system. IEEE Std 7000™-2021 further elaborates how values can be elicited and explicitly considered as part of the design process of any product or service. In the context of value-sensitive development the public discourse should be performed throughout the lifecycle of an automated vehicle. Initially, the public debate serves as a basis for the stakeholder related analysis of values and, if possible, the elicitation of context specific requirements. As a prerequisite of releasing automated vehicles will be the definition of acceptance criteria, which will likely be documented as part of standards or laws, this explicit exchange of values has to be performed before the release of the technology. Furthermore, acceptance criteria are volatile. They can change due to specific events as it was observable for nuclear power plants e.g. in Germany, and might change over time as automated vehicle technology matures.

Throughout the design and operation of the automated vehicle, stakeholder values build the reference for validation of the implemented behavior. In order to integrate stakeholder values into the development process, Graubohm et al. [60] present a value-sensitive design [61] inspired approach for the concept phase of an automated vehicle. The iterative nature of value-sensitive design is explicitly built into the process model and facilitates the adjustment of value trade-offs. How these trade-offs are performed will be highly dependent on the defined acceptance criteria and possibly even company culture. Hence, a traceable and transparent documentation of assumptions and design decisions is required in order to be able to argue for sufficient safety. A framework for performing value trade-offs is for example described in the IEEE Std 7000™-2021 Standard.

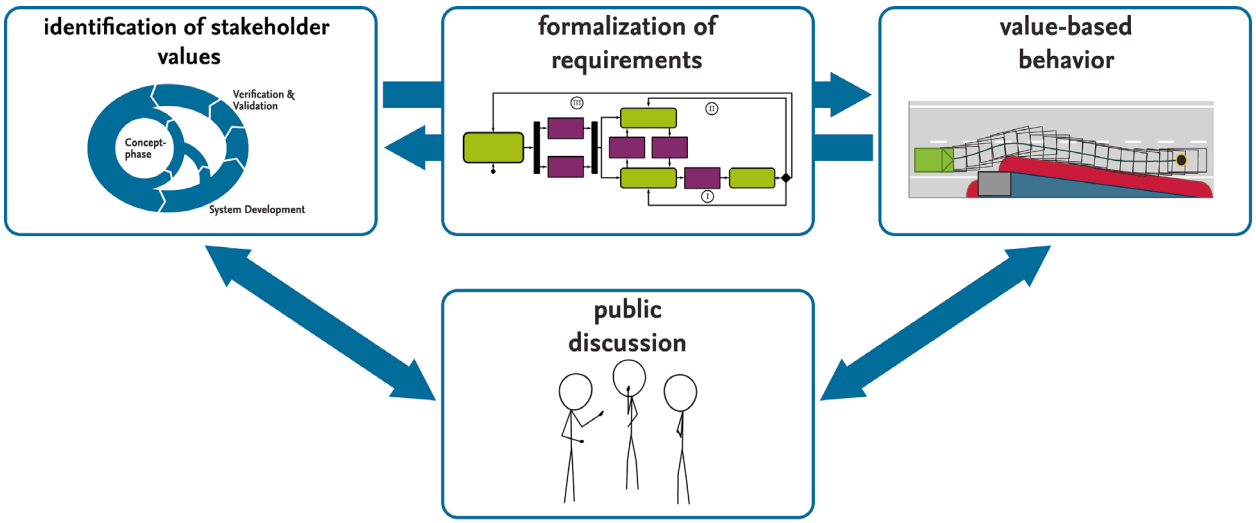

Fig. 40.5 General framework for value-sensitive development of automated vehicles

Since values tend to be abstract and difficult to generally weigh against each other without a given context [62, 63], Salem et al. [64] and Beck et al. [65] introduce the semantic norm behavior analysis and phenomenon signal model as methods to contribute to a context specific behavior specification. The authors use a scenario-based analysis to define concepts and rules for a given scenario and finally validate the formally inferred target behavior against the elicited values, or in their specific approach, behavioral norms. With respect to weighing values, Millar et al. [63] perform a context-specific assessment of values and necessary trade-offs within stakeholder workshops. Hence, the elicitation of values, trade-offs and their respective integration in a formal behavior specification could for example be performed by combining the aforementioned approaches. The scalability of such a proceeding, especially in the context of assuring safety of automated vehicles, remains to be analyzed.

Another aspect of value-sensitive development is the implementation of the specified vehicle behavior. For the tactical formulation of weights and constraints within a Model Predictive Control framework, Nolte et al. [66] present an architectural approach that enables a traceable integration of target behavior in the design of an automated vehicle. In order to assess this implemented behavior with regard to reasonable risk—or violation of stakeholder values—a transparent communication of the safety argument and the included value trade-offs will be required. Finally, all of the aforementioned approaches can contribute to an overarching value-sensitive development framework.

While giving a brief overview of a value-sensitive perspective on safety, we first elaborated a respective meaning and use of the term safety and subsequently discussed possible challenges in the implementation of a value-sensitive development framework. These challenges include aspects like an uncertain scalability of the overall approach as values and especially value trade-offs are context-dependent. Furthermore, integrating the public in debates on acceptance criteria regarding values can lead to development uncertainties for companies as they require very rigorous and obligatory agreements especially regarding safety. An additional obstacle with respect to integrating explicit value trade-offs in a notion of safety is the necessary transparency in a safety argument. Since there is not only technical but especially legal and ethical considerations included in such trade-offs it is a foreseeable challenge for the communication of companies developing automated vehicles.

However, the uncertainty involved in assuring safety of automated vehicles will inherently result in scenarios in which developers will be required to argue for specific design decisions and assumptions that lead to the observed behavior (cf. example in ▶ Sect. 40.2.1.2). Thus, a value-sensitive development approach can extend an argumentation for safe behavior (especially in edge cases).

40.2.1.4 Summary

In this section we discussed the meaning and use of the term safety in three engineering frameworks for designing safe automated vehicles. First, we showed how established frameworks for assuring functional safety and SOTIF take an accepted risk criterion that is set as a design goal, which are expressed as safety integrity levels in the case of functional safety. Current risk acceptance criteria used or those that could be used in the future focus on defining a level of risk in terms of the probability of physical harm that is considered reasonable. Thus, the assurance of a certain safety integ-

rity, which is based on a classification of the probability and severity of physical harm associated with specific hazards, is sufficient to classify a system as 'safe' in the scope of functional safety.

Next, we gave examples of formal methods for safety assurance and showed what they mean by "safety". Based on the approach of proving safety mathematically, there are multiple assumptions being made by for example the frameworks of Safety Force Fields [54], Reachable Set Analysis [55] and Responsibility Sensitive Safety [53]. These assumptions need to be agreed upon as being reasonable in order to deem a system as being safe. However, we elaborated that such approaches do neither account for the consideration of risk acceptance criteria nor an explicit documentation of trade-offs that are involved in specifying safe behavior of an automated vehicle.

Finally, we introduced the framework of value-sensitive development, where safety is a state in which stakeholder values are being at reasonable risk. By adding further categories of risk into the notion of safety that extend the focus on physical wellbeing to additional values such as mobility, a value-sensitive development framework acknowledges the importance of value trade-offs when specifying safe behavior. We also highlighted challenges like uncertainty regarding scalability, transparent risk communication, and the challenge of agreeing on acceptance criteria while accounting for multiple stakeholder values.

40.2.2 How the Term 'Safety' is Used by Other Stakeholders

In the previous section we analyzed different uses of 'safety' in the automotive safety engineering community, focusing on 'safety' as a concept that acknowledges the existence of residual risk. According to the analysis so far, the risks included in each definition can differ from one definition to another, leading to a different meaning of 'safety' in each case. We extend this analysis here and survey several credible and influential sources to gain an understanding of how the term 'safety' is used in relation to automated vehicles among four other stakeholder groups: regulators and legislators, corporations, advocacy groups and public media. Note that we include examples of statements from these groups and do not claim to comprehensively represent all relevant perspectives. We organize and analyze how the term 'safety' is used by a stakeholder group (at least partially) because the perspective may have an impact on the type of risks included in a safety claim about automated vehicles.

What we find in our analysis in this section is that the definitions of 'safety' are sometimes stated explicitly, but are also often implied, along with the associated risks. Sorting out what is meant in any safety claim related to automated vehicles thus requires some further guesswork or analysis. What we also find in our analysis is that safety is talked about differently when comparing statements from one stakeholder group to another group. However, while safety statements differ within a single group, there are also several similarities in how safety is talked about within a group.

40.2.2.1 Legislators and Regulators

In 1997, the Road Traffic Safety Bill founded on Vision Zero was passed by a large majority in the Swedish parliament. The Vision Zero policy sets out aspirational goals in that "eventually no one will be killed or seriously injured within the road transport system", while "long-term disabling injuries, and non-injury accidents are more or less outside the scope of the Vision" [67]. In this goal, The Vision Zero policy is making a safety claim, and is explicit about the associated risks that are included and excluded in the claim. We can understand that the policy is making a safety claim that accounts for the risk of any serious injury or lethal accidents and neglects risks associated with less serious or non-injury accidents. We can also understand that any probability of occurrence of a serious injury or lethal accident is unreasonable in Vision Zero's safety claim.

More recently in 2020, a project and report was undertaken by The Expert Panel of Connected and Autonomous Vehicles and Shared Mobility, which was assembled by The Canadian Council of Academies upon request by the federal institution that leads the Industry, Science and Economic Development portfolio in Canada. In the report, The Expert Panel presents road safety policy recommendations based on the statement that "fully autonomous vehicles could further reduce injuries and fatalities by eliminating human error in driving, which is a factor in over 90% of collisions" [68]. This statement in part draws on a widely cited statistic from the United States National Highway Traffic Safety Administration [6], which is further elaborated on NHTSA's website page discussing the safety benefits of AVs: "Automated vehicles' potential to save lives and reduce injuries is rooted in one critical and tragic fact: 94% of serious crashes are due to human error [in driving]. Automated vehicles have the potential to remove human error [in driving] from the crash equation" [69]. In this statement from NHTSA, which informs policy recommendations and statements such as those made by The Expert Panel, NHTSA is making a safety claim that is different from Vision Zero's claim. What we can understand from NHTSA's safety claim is that it addresses risks of serious injury or lethal accidents caused by human error, whereas Vision Zero does not focus on whether accidents are caused by human error. We can

also understand that NHTSA's safety claim considers severe or lethal accidents with a probability of occurrence that is lower than a referenced probability for manually driven vehicles, whereas Vision Zero's claim addresses severe or lethal accidents with no probability of occurring.

NHTSA also mentions risks of injuries and crashes in its mission to "Save lives, prevent injuries, reduce vehicle-related crashes". However, in other major statements by NHTSA, it is not clear what NHTSA means by the term safety. For example, during a keynote speech at the Automated Vehicles Symposium, NHTSA's Deputy Administrator stated that the development of automated vehicles will "promote safety" without further clarifying the term safety [70]. Distinguishing NHTSA's safety claim from others' safety claims is thus not straightforward in this case.

In Germany the traffic law requires automated vehicles to transition into a minimal risk condition in order to ensure the highest possible traffic safety for passengers, other traffic participants, and third parties in a given situation. However, it is not further specified, which risks should be minimized exactly. Thus, we assume that, based on the terminology used in several paragraphs of the German traffic law (e.g. StVG sec. 40.1e), the considered risks account for physical injuries due to accidents. Further challenges in the context of minimal risk conditions become evident by comparing the use of the term 'safety' between a legal and a technical perspective. The ISO/TR 4804:2020 describes a minimal risk condition as a condition of an automated vehicle that can be transitioned into by executing a minimal risk maneuver. Given the potential mismatch of a condition with an estimated minimal risk and the condition actually associated with a minimal risk, based on ex post data analysis [71], the definition of appropriate minimal risk maneuvers becomes even more challenging. Thus, fulfilling the legal requirement of transitioning into a minimal risk condition requires a clarification of how the definition of such a condition would be assessed retrospectively.

40.2.2.2 Corporate Statements

Corporations from the automotive industry have made several claims about the expected safety of their vehicles. For example, Tesla's Vehicle Safety Report states that "That's why Tesla vehicles are engineered to be the safest cars in the world… [and] have achieved the lowest overall probability of injury" [72]. From this statement, Tesla is making a safety claim which includes the explicit risk of road injuries. We can also guess implicitly that Tesla's safety claim is associated with the probability of a road injury occurring that is lower than a referenced probability of occurrence.

Other corporations such as Waymo are more explicit in their Safety Report about what they mean by the term safety. In the Glossary of Waymo's Safety Report, safety is defined as "Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment" [73]. The risks associated with Waymo's safety claim are explicit and account for the risk of death or injury which are similar to the risks addressed in the previous safety claims that we have reviewed so far. Furthermore, we can understand that from Waymo's use of the word 'freedom' in their safety claim, their claim includes risks with no probability of occurrence, which is similar to Vision Zero's focus on the absence of risk in their safety claim. The risks associated with Waymo's safety claim differs from the previous claims in that the risks are associated with "conditions" (such as accidents) that can cause death or injury, thus capturing a broader list of risks from the previous claims which are associated with just accidents. The risks associated with Waymo's safety claim also include other risks than physical injuries, such as occupational illness (possibly referring to health issues beyond injuries), damage to equipment or property (possibly from the result of a collision between vehicles or between vehicles and the surrounding infrastructure), and damage to the environment (such as perhaps pollution caused by the accumulation of vehicle exhaust systems on roads). Overall, we can understand that Waymo is including further risks in their safety claim associated with AVs compared to the previous safety claims we have reviewed so far.

Volvo Group on the other hand is not explicit about what they mean by the term safety. For example, in a column on Volvo's website, The Safety Director at Volvo Group wrote about whether automated vehicles will improve safety without any further clarification of what the term safety means [74]. Distinguishing Volvo Group's safety claim from others' safety claims is thus not straightforward.

40.2.2.3 Advocacy Groups

Advocacy groups such as Mothers Against Drunk Driving (MADD) Canada support the development of automated vehicles to improve road safety. The President of MADD Canada communicated this support stating that automated vehicles will "eventually eliminate impaired driving completely" and that "many lives will be saved due to the reduction in impaired driving and driver error" [75]. The implicit risk included in this safety claim is likely that impaired driving can lead to severe or lethal accidents. We can also understand that the risk of impaired driving is associated with a probability of occurrence that is reduced compared to a referenced probability. Similar to NHTSA's safety claim above, the safety claim from Mothers Against Drunk Driving includes the risk of an accident associated with

driver error, given that impaired driving can be understood as a type of human error in driving.

Other organizations such as End Distracted Driving (EndDD) have talked about the potential of automated vehicle features to improve road safety. In an article on their website, the organization makes the safety claim that "autonomous features will still be there to protect all of us […] against distracted driving" [76]. In a second article, the organization also states that automated driving features could "improve their vehicles' ability to detect other cars in the road" as well as "protect [distracted] pedestrians from themselves" and "spot pedestrians and cyclists from a few yards away, predict their trajectory, sound an alarm and, if necessary, activate the brakes if you get too close" [77]. The implicit risks associated with the safety statement in the first article is likely that, similar to the statements reviewed above from Mothers Against Drunk Driving about impaired driving, distracted driving can lead to severe or lethal accidents. Furthermore, the safety statement in the second article seems to include an implicit risk associated with the limits of human driver reaction time to compensate for other distracted drivers or distracted pedestrians.

Advocacy groups such as the National Safety Council (NSC) also spoke about distracted driving in their statements about the potential of automated vehicles to improve road safety. In a Testimony before the Senate Appropriations Committee Transportation, Housing and Urban Development, and Related Agencies Subcommittee, Deborah A.P. Hersman, the former President and CEO of NSC and the former Chief Safety Officer at Waymo, stated that automated driving features would "allow a vehicle to communicate with a red light to compensate for a fatigued driver, stop a car to prevent a collision with a pedestrian if a driver fails to detect him or her, and prevent or mitigate collisions between vehicles" [78]. From this statement, we can understand that the risks associated with safety include risks of collisions (either with a pedestrian or with another vehicle) associated with distracted or fatigued driving.

The organization Let's Talk about Autonomous Driving (LTAD) further represents a diverse set of communities which include Mothers Against Drunk Driving and the National Safety Council, as well as the Foundation for Senior Living, the Foundation for Blind Children, and Students Against Destructive Decisions. On LTAD's website, the organization states that these communities share the belief that "autonomous driving cars can save lives" [79]. In an article on LTAD's website about the links between automated vehicles and safety/stress, the organization further talks about the potential safety benefits of automated vehicles "to reducing stress, road rage, and reckless driving" [80]. The organization further cites evidence that suggests that the stress of driving in traffic can harm a person's psychic well-being. Thus the organization's safety statement about stress, road rage, and reckless driving addresses risks of psychic harms.

40.2.2.4 Public Media

The public media such as Wired and Canadian Broadcasting Corporation (CBC) are releasing stories about safety concerns associated with automated vehicles. In a story released by Wired about the links between automated vehicles and safety/health concerns, Peter Norton, a transportation historian at the University of Virginia says "Why do we care about safety? Because we care about human health" [81]. Norton also warns that the automated future will result in less walking, more sprawl, and—without dedication to electric propulsion—a huge spike in pollution. The article goes on to cite evidence that suggests sedentary behavior and pollution associated with driving are linked to health risks of cancer and heart disease. The safety concerns associated with AVs in this article differs from the previous statements we have reviewed, which tend to focus on the safety benefits of AVs in removing human drivers from the driving task. While human driven cars kill people in crashes, the article states that cars driven on roads also lead to sedentary behavior and pollution, and that depending on how AVs are designed there could be an increase in cars being driven on roads. Overall, from the statements made in this article we can understand that the risks associated with the safety of AVs include risks of cancer and heart disease due to sedentary behavior and pollution.

Another story released by the CBC talks about the links between automated vehicles and safety/security concerns, where the term security tends to be related to the term safety in that security addresses the risk of harm due to intentional criminal acts. In the CBC article, Sebastian Fischmeister, a Professor in the Department of Electrical and Computer Engineering at Waterloo University in Canada, says "Now that all these vehicles are connected, you have the risk of someone sitting somewhere and through the internet and through connectivity reaching out to all vehicles of a particular manufacturer," and that "That's why safety and security are the key challenges" [82]. Here we understand that Fischmeister is including risks of hackers in their concept of safety, suggesting that automated vehicles are increasingly being built with more software components which can be hacked and controlled. A further implied risk made in Fischmeister's statement may be the risk of loss of control of the vehicle to terrorist hackers, and a hacker's potential ability to control hundreds or thousands of vehicles all at once from the comfort of their home.

40.2.2.5 Summary

Focusing on "safety" as a concept that includes a bundle of risks, we break down safety claims associated with automated vehicles that are made by different stakeholders. We find that stakeholders such as legislators and regulators, corporations, advocacy groups, and public media talk about safety differently given that, when the risks were made explicit, there are different risks addressed among stakeholder safety claims. We find that legislators and regulators tend to talk about the risk of serious injury or fatal accidents when talking about automated vehicle safety, where their claims differ depending on whether their claims account for accidents in general or accidents associated with driver error, and depending on whether their claims consider accidents with a lower probability of occurring (compared to a referenced probability) or accidents with no probability of occurring. In contrast, we find advocacy groups tend to talk about the risk of accidents associated with impaired, fatigued, or distracted driving. Notably we also found an organization of advocacy groups that spoke about the risks of psychic harms and stress associated with driving in their claim about automated vehicle safety. Furthermore, we find that some corporations cover a broader list of explicit risks in their safety claims, such as occupational illness, damage of equipment or property, and damage to the environment caused by "conditions", while other vehicle corporations were not explicit about the risks associated with their safety claims, requiring further guesswork or analysis to distinguish their claims from others. Finally, we find that the public media covered broader risks associated with automated vehicle safety concerns, including the risk of heart disease and cancer due to sedentary behavior and pollution, as well as the risk of loss of control due to hackers taking over control of the vehicles remotely.

Based on the uses of the term 'safety' by multiple stakeholders and our analysis of safety engineering frameworks in ► Sect. 40.2.1 we summarize implications with respect to our hypothesis of a diverging meaning and use of the term 'safety'.

40.2.3 Conclusions About the Meanings and Uses of the Term 'Safety'

In the previous section we analyzed how the term 'safety' is used in selected safety engineering frameworks and how other stakeholders use the term. We focused our analysis on the relation between different kinds of risk being considered when raising safety claims or concerns.

For the stakeholder group of engineers we elaborated how safety is traditionally handled as the absence of unreasonable physical harm and how existing standards provide guidelines on how to assure safety of automated vehicles within this defined understanding of related risks. We also pointed out challenges that emerge from the shift from assisted to automated driving with regard to safety measures and quantitative risk assessments. These challenges include for example the absence of the measure of controllability, the issue of adhering to a positive risk balance given differently severe accidents, and more generally the determination of quantitative risk acceptance criteria for the new technology of automated vehicles (SAE Level 3+).

Additionally, we showed examples of approaches for safety engineering from corporate and academic research that use the term 'safety' differently. On the one hand there are approaches that address safety as a formally provable property of an automated vehicle. Whereas formal proofs provide (economically) scalable frameworks, they require fundamental assumptions on reasonable parameters chosen that implicitly might lead to safe systems. However, the relation between a safe vehicle and the risks that are associated with its operation remain implicit and hence unclear.

On the other hand we presented an approach for value-sensitive development in order to give an example of how safety can also be understood as a concept that accounts for stakeholder objectives explicitly. Based on such an explicit consideration of stakeholder values, the risks that are associated with an automated vehicle can be evaluated and explicitly included in the design of these systems.

Other stakeholder groups use the term 'safety' and are not always explicit about the risks they include in their understanding of a safe automated vehicle. Furthermore, the statements we collected as part of this book chapter do not mean to represent a consistent view within each stakeholder group. However, we support our claim that safety is used differently across different stakeholders. In addition to the kinds of risk that are included, which range from physical harm to for example damage of equipment or property, we elaborated that there is a plethora of further considerations that go into the classification of a 'safe' automated vehicle. For example, one question could be if automated vehicles lead to less accidents in road traffic than human drivers in general or if they perform better than impaired drivers. This question becomes even more demanding by comparing automated vehicles with performant and attentive drivers.

Given the provided evidence for different meanings and uses of the term 'safety', in the following section we provide a set of questions about aspects of risk that could be asked when facing safety claims or concerns from different stakeholders. These questions can be used to facilitate an explicit and more detailed distinction between meanings of safety and associated risks and can further support an open discussion among

stakeholder groups and disciplines about approaches for handling safety and risk communication for AVs.

40.3 Questions About the Risks Associated with the Term 'Safety'

In the previous section we gave examples of different meanings and uses of the term 'safety'. We found that different stakeholders include different risks when they raise concerns about the safety of automated vehicles or address these concerns with respective claims in a safety engineering framework.

In this section we pose questions about the risks associated with the term 'safety' in order to breakdown the risks into aspects that can be compared. These aspects of risk that we found are as follows: category, scope, target, time, reference, and stakeholder-dependence. The questions we pose to consider these aspects of risk can facilitate future discussions among stakeholder groups and disciplines about how safety statements and associated risks differ, and about approaches for handling safety communication for automated vehicles.

The first question that can be raised to break down a safety concern or claim is:

Which objective is at risk? (category)

This aspect specifies the category of the considered risks. Given the definitions of functional safety (ISO 26262:2018) and safety of the intended functionality (ISO 21448:2022), we elaborated that the objective at risk in these meanings of safety is the physical well-being of persons involved in an operational situation. Statements from Waymo extend this meaning to further objectives such as occupational illness and damage of property. These extensions of the risks to be considered when using the term 'safety' can be generalized as defined for risk management in the ISO 31000:2018. Here, any objective under uncertainty is included as a category of risk that can be managed.

In order to explicitly state the scope when using the term 'safety', one can raise the question:

What scope is considered with the risks? (scope)

Again starting from established safety engineering frameworks for functional safety and SOTIF, hazards and their associated risks are explicitly considered at the vehicle level [22, 26]. Hence, the system boundary is drawn between the vehicle and its immediate environment in an operational situation or scenario [52]. Statements in the public press like those from Peter Norton in Wired suggest to consider risks on a societal level in order to include health risks caused by for example less walking due to automated driving. Based on this distinction, it becomes clear that different risks can be considered depending on the system boundary or the scope.

Furthermore, the entity which is at risk can be made explicit by asking the question:

Who is at risk? (target)

The answer to this question also depends on the scope at which risks are included in a meaning of safety. For example in a scenario where only the immediate vicinity of the vehicle is considered, passengers and other road users could be considered as affected by different risks. In the ISO 26262:2018 possible targets of harm explicitly defined as "the driver, passengers or persons in the vicinity of the vehicle's exterior". With regard to a larger scope there could also be companies at risk or specific parts of society that do not necessarily interact with automated vehicles directly. When introducing RSS, Shalev-Shwartz et al. [53] write: "The second area of risk lies with lack of scalability." We attribute the kind of entities concerned with such a risk as being companies for which scalability is an objective.

In our analysis we elaborated that there can be differences between considering risks as an outcome of deploying the system and estimating possible risks and defining design goals. Hence, we ask the question:

Are the risks judged retrospectively or prospectively? (time)

The goal of achieving a positive risk balance, which was raised as a requirements by the German ethics commission, and translating that into design goals for an unreasonable amount of risk was for example discussed by Favaró [83]. This process has plenty of challenges as elaborated by Favaró. One of these challenges is changing the perspective towards the deployment of automated vehicles regarding the aspect of time. A positive risk balance could for example be interpreted as the reduction of fatalities in road traffic for each year individually after the technology is released. Also one could tolerate a relative increase in fatalities in the first year but manage the balance with a stepwise market penetration as Wachenfeld et al. [15] suggested and as it is practiced in aviation domain [17].

Even though we focus our analysis on the meanings and uses of the term 'safety' and do not attempt to define an amount of risk that could be reasonable for automated vehicles, the following question helps identifying different kinds of risks:

With respect to which reference does the system pose risks? (reference)

Whether or not an automated vehicle is reasonably safe requires a clarification of how "reasonable" is defined.

Quantitative risk acceptance criteria take different reference values of risks below which an automated vehicle could be considered reasonably safe. These can for example be the probability of being struck by a lightning or of being involved in an accident with a conventional road vehicle. Advocacy groups like Mothers Against Drunk Driving (MADD) or End Distracted Driving (EndDD) define the goal of reducing accidents due to different sorts of impaired driving. Also, statements from for example NHTSA [69] suggest to compare automated vehicles to human drivers and assess the impact of mitigating human error for the safety of road traffic. A more rigorous reference is given by Vision Zero with the goal of completely eliminating fatalities and severe accidents in road traffic. Waymo's definition of safety—comparable to Vision Zero—suggests a complete freedom of harmful conditions. Therefore, in both cases the absence of all risk emerging from automated driving could also be understood as a possible reference to evaluate the safety of automated vehicles.

As we compared perspectives from different stakeholders on what constitutes the safety of a system with respect to the included risks we found that there are risks that only apply to some group of stakeholders and those that can be applied to all of them. The following question helps to make this difference explicit:

> Are the risks based on a perspective from a group of stakeholders or do they apply to all stakeholders? (stakeholder-dependence)

For example an aspect of risk that was only implicitly addressed in this analysis so far is concerned with factors of risk perception. Statements from Waymo considering occupational illness suggest that perceived safety can play a major role for the acceptance of automated driving. The presented framework for value-sensitive development can facilitate including such considerations by accounting for risk perception in the process of weighing stakeholder values. On the other hand, established safety engineering frameworks and formal safety assurance methods rely on mathematical metrics to assess different kinds of risk. These metrics are especially necessary to find a common ground for different stakeholder groups with otherwise very heterogeneous subjective perceptions of risk. However, this does not imply that a seemingly more objective "expert-based risk assessment" is superior to aspects of for example public risk perception as also pointed out by Slovic [42].

40.4 Limitations and Recommendations

In the previous chapter we analyzed different stakeholder perspectives on the safety of automated vehicles and how they use the term 'safety'. We elaborated that there are different kinds of risk being considered by those stakeholder groups. Our analysis is limited in so far that we can neither claim to have considered all relevant stakeholder groups nor that those included are being comprehensively represented. We deem this limitation to be acceptable as it does not compromise our claim that there are different meanings and uses of the term 'safety'.

In order to support future research on this issue we give recommendations on how this initial analysis could be complemented. From an engineering perspective we only included examples of automotive safety engineering frameworks. This selection and thereby also our classification of three kinds of frameworks could be extended by analyzing further standardization work such as the SAE J3016:2021, SAE J3018:2019, ISO/TR 4804:2020, or IEEE 7000-2021. Additionally, further research in the field of automated vehicle safety could be included such as work from Koopman et al. [84], Wachenfeld [12], Fraade-Blanar et al. [85], or Stolte et al. [86]. We also did not include perspectives of stakeholder groups from other parts for the world such as Asia. Thus, there are other influential and reliable sources whose safety claims and associated risks could be included in future analyses. Regarding experience from other domains such as aviation or rail, we found differences with respect to the issue of an open traffic context and the coping mechanisms that can be put in place in order to assure safety. For example the existence of air traffic control enables different safety argumentation strategies compared automated driving. Hence, we did not focus on notions of safety in these domains but we acknowledge the potential of including considerations about safety from these and further fields (like robotics) in the automotive domain.

Generally, we follow the recommendation from Homann [87, 88], who advocated a public discourse of risks in the context of driver assistance systems in 2002, and support his approach in the domain of automated vehicles. Given Homann's argumentation, we further recommend to intensify efforts of an open discourse of risks associated with automated driving *before* the release of respective systems. This becomes increasingly important as first automated driving systems with SAE Level 3 have recently been certified in Germany [89].

Burton et al. [90] present an analysis of potential gaps in safety assurance from an interdisciplinary perspective and offer suggestions for the practical realization of their findings. Thus, Burton et al. [90] recommend to thoroughly define the operational design domain of the automated driving system in order to provide a basis of the system specification. Furthermore, an iterative refinement of safety assurance activities in parallel to the system design are required in order to identify safety-critical scenarios and system limitations as soon as possible and define necessary measures

like limiting the operational design domain or adapting the system design (cf. SOTIF [22]). Finally, Burton et al. [90] suggest that in liability cases, there should be an interdisciplinary evaluation where assumptions and decisions in the design process are assessed.

40.5 Summary and Outlook

With 'safety' and 'risk' being two terms of major importance in the field of automated driving this chapter provides a selection of meanings and uses of these terms that are found either in safety engineering frameworks or in statements of different stakeholder groups. The analysis showed that there are fundamental differences in the kinds of risks that are addressed in various safety claims and concerns.

In section ▶ Sect. 40.1 we argued that the shift from assisted to automated driving implies new challenges for safety assurance. We discussed the role of uncertainty in the design of automated vehicles and the participation in road traffic. As a result of these uncertainties in combination with stakeholder objectives, we referred to the concept of inherent risk. Based on this motivation of why safety and risk as concepts should be considered differently for automated instead of assisted driving, we raised the concern that there are different meanings and uses of these terms among stakeholders and this might lead to mismatched expectations of the safety of automated vehicles.

Hence, we gave examples of safety engineering frameworks that are either used in the automotive industry or proposed by corporate or academic research groups. Additionally, we gathered statements from different stakeholder groups. Based on an analysis of these selected frameworks and statements, we concluded that there are in fact different meanings and uses of the term 'safety' and these differences highly depend on the kinds of risk that are considered (▶ Sect. 40.2).

As a refinement of the claim of mismatched meanings and uses of the term 'safety', we provided a set of questions to further break down a safety concern into the included aspects of risk (▶ Sect. 40.3). These questions help to clarify and compare the details of the different risks associated with the meanings and uses of the term 'safety' that we found in our analysis of safety engineering frameworks and stakeholder concerns.

In order to support future research we gave recommendations to extend this initial analysis by including further work addressing safety terminology of automated vehicles. An additional challenge, which was not part of this chapter, lies in answering the question of "How safe is safe enough?" This question can be raised on the societal level but also on the vehicle level. With automated vehicles being required to be capable of transitioning into a minimal risk condition for example by the German traffic law, clarity about aspects of risk that should be considered and the applied metrics and criteria for automated driving is important as well. As an answer to this question depends on a common understanding of different meanings and uses of the term 'safety' we suggest a more transparent communication of assumptions that are made when raising safety claims or concerns.

Lastly, in a context of rapid technological change with urgent societal and environmental implications, it is essential that trust remains the bedrock of societies, communities, economies, and sustainable development [91]. While the focus of further research has been on addressing issues which affect the release and acceptance of automated vehicles, further research is also needed to realize, identify, and address related issues which affect the perceived or actual trustworthiness of automated vehicles and associated actors and processes, which can lead to lower trust among stakeholders such as companies or developers and the public.

Acknowledgements For their financial support of this work we thank the German "Federal Ministry of Economic Affairs and Climate Action" in the context of the project "Verification and Validation Methods for Automated Vehicles Level 4 and 5", the "Daimler and Benz Foundation" in the context of the project "Value-based decision-making", and the German "Federal Ministry of Education and Research" for funding the project "Automatisiertes Fahren im Mischverkehr". We further thank Till Menzel and Torben Stolte for their valuable feedback on this chapter. Special thanks go to Hermann Winner, who has fundamentally contributed to the discussions that where part of developing this work.

References

1. Grunwald, A.: Societal risk constellations for autonomous driving. Analysis, historical context and assessment. In: Autonomous Driving, pp. 641–663. Springer (2016)
2. Tsugawa, S., Jeschke, S., Shladover, S.E.: A review of truck platooning projects for energy savings. IEEE Trans. Intell. Veh. **1**(1), 68–77 (2016). ▶ https://doi.org/10.1109/TIV.2016.2577499
3. Shladover, S.E.: PATH at 20—history and major milestones. IEEE Trans. Intell. Transp. Syst. **8**(4), 584–592 (2007). ▶ https://doi.org/10.1109/TITS.2007.903052
4. Dickmanns, E.D., Zapp, A.: Autonomous high speed road vehicle guidance by computer vision 1. In: IFAC Proceedings Volumes, 10th Triennial IFAC Congress on Automatic Control—1987 Volume IV, Munich, Germany, 27–31 July 20, pp. 221–226 (1987). ▶ https://doi.org/10.1016/S1474-6670(17)55320-3
5. Smith, B.W.: Human error as a cause of vehicle crashes. The Center for Internet and Society, Stanford Law School (2013). ▶ http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes. Accessed 13 Dec 2021

6. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey (No. DOT HS 812 115). NHTSA's National Center for Statistics and Analysis, United States, DC (2015)
7. Bagschik, G.: Personal communication (2016)
8. Maurer, M., Gerdes, J.C., Lenz, B., Winner, H.: (eds.): Autonomous Driving. Springer Berlin Heidelberg, Berlin, Heidelberg (2016). ► https://doi.org/10.1007/978-3-662-48847-8
9. Schnieder, E., Schnieder, L.: Verkehrssicherheit. Springer, Berlin, Heidelberg (2013). ► https://doi.org/10.1007/978-3-540-71033-2
10. Reschka, A., Bagschik, G., Maurer, M.: Towards a system-wide functional safety concept for automated road vehicles. In: Automotive Systems Engineering II, pp. 123–145. Springer (2018)
11. Bagschik, G., Reschka, A., Stolte, T., Maurer, M.: Identification of Potential Hazardous Events for an Unmanned Protective Vehicle (2018). arXiv preprint ► arXiv:1804.08728
12. Wachenfeld, W.H.K.: How Stochastic Can Help to Introduce Automated Driving. Technische Universität Darmstadt, Germany, Darmstadt (2017)
13. Stolte, T., Bagschik, G., Reschka, A., Maurer, M.: Hazard analysis and risk assessment for an automated unmanned protective vehicle. In: 2017 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 1848–1855 (2017). ► https://doi.org/10.1109/IVS.2017.7995974
14. Kühn, M., Hannawald, L.: Verkehrssicherheit und Potenziale von Fahrerassistenzsystemen. Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort (2015)
15. Wachenfeld, W., Winner, H.: The new role of road testing for the safety validation of automated vehicles. In: Automated Driving: Safer and More Efficient Future Driving (2017)
16. Junietz, P., Bonakdar, F., Klamann, B., Winner, H.: Criticality metric for the safety validation of automated driving using model predictive trajectory optimization. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Presented at the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 60–65 (2018). ► https://doi.org/10.1109/ITSC.2018.8569326
17. Junietz, P.M.: Microscopic and Macroscopic Risk Metrics for the Safety Validation of Automated Driving (Dissertation). Technische Universität, Darmstadt (2019). ► https://doi.org/10.25534/tuprints-00009282
18. Gasser, T.M., Schmidt, E.A., Bengler, K., Chiellino, U., Diederichs, F., Eckstein, L., Flemisch, F., Fraedrich, E., Fuchs, E., Gustke, M., Hoyer, R., Hüttinger, M., Jipp, M., Köster, F., Kühn, M., Lenz, B., Lotz-Keens, C., Maurer, M., Meurer, M., Meuresch, S., Müller, N., Reitter, C., Reschka, A., Riegelhuth, G., Ritter, J., Siedersberger, K.-H., Stankowitz, W., Trimpop, R., Zeeb, E.: Bericht zum Forschungsbedarf. Runder Tisch Automatisiertes Fahren—AG Forschung. Federal Ministry of Transport and Digital Infrastructure (2015)
19. Di Fabio, U., Broy, M., Jungo Brüngger, R., Eichhorn, U., Grunwald, A., Heckmann, D., Hilgendorf, E., Kagermann, H., Losinger, A., Lutz-Bachmann, M., Lütge, C., Markl, A., Müller, K., Nehm, K.: Ethics Commission—Automated and Connected Driving. Federal Ministry of Transport and Digital Infrastructure (2017)
20. SAE: J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. SAE International (2021)
21. Safety: In Merriam-Webster.com Dictionary (n.d.). ► https://www.merriam-webster.com/dictionary/safety. Accessed 13 Oct 2021
22. International Organization for Standardization: ISO 21448 Road vehicles—Safety of the intended functionality (2022)
23. Shladover, S.E.: Framework for safety assurance of automated driving systems: unresolved issues. In: Gehalten auf der 27th ITS World Congress, Hamburg, Germany (2021)
24. Maurer, M.: Hochautomatisiertes und vollautomatisiertes Fahren. In: Deutscher Verkehrsgerichtstag—Deutsche Akademie für Verkehrswissenschaft (Ed.), 56. Deutscher Verkehrsgerichtstag 2018. Presented at the 56. Deutscher Verkehrsgerichtstag 2018, pp. 43–57. Luchterhand Verlag (2018)
25. Nolte, M., Ernst, S., Richelmann, J., Maurer, M.: Representing the unknown—impact of uncertainty on the interaction between decision making and trajectory generation. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Presented at the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2412–2418 (2018). ► https://doi.org/10.1109/ITSC.2018.8569490
26. International Organization for Standardization: ISO 26262 Road Vehicles—Functional Safety, Part 1: Vocabulary (2018)
27. International Electrotechnical Commission: IEC 61508-1, Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems—Part 1: General Requirements (2010)
28. Deutsches Institut für Normung: DIN EN ISO 12100, Sicherheit von Maschinen—Allgemeine Gestaltungsleitsätze—Risikobeurteilung und Risikominderung (2011)
29. International Organization for Standardization, International Electrotechnical Commission: ISO/IEC Guide 51, Safety Aspects—Guidelines for Their Inclusion in Standards (1999)
30. Koch, K.-R.: Einführung in die Bayes-Statistik. Springer Berlin Heidelberg, Berlin, Heidelberg (2000). ► https://doi.org/10.1007/978-3-642-56970-8
31. International Organization for Standardization: ISO 26262 Road Vehicles—Functional Safety—Part 3: Concept Phase (2018)
32. Society of Automotive Engineers: SAE J2980—Considerations for ISO 26262 ASIL Hazard Classification (2018)
33. Douglas, M., Wildavsky, A.: How can we know the risks we face? Why risk selection is a social process. Risk Anal. **2**, 49–58 (1982). ► https://doi.org/10.1111/j.1539-6924.1982.tb01365.x
34. Crouch, E.A.C., Wilson, R.: Risk/Benefit Analysis. Ballinger Publishing Company, Cambridge, MA (1982)
35. Gibson, S.B.: The use of quantitative risk criteria in hazard analysis. J. Occup. Accid. **1**, 85–94 (1976). ► https://doi.org/10.1016/0376-6349(76)90009-2
36. Kinchin, G.H.: Design criteria, concepts and features important to safety and licensing. ANS, In: Proceedings on the International Meeting on Fast Reactor Safety Technology, WA, United States (1979)
37. Kletz, T.A.: Hazard analysis, its application to risks to the public at large. Reliab. Eng. **3**, 325–338 (1978)
38. Starr, C.: Social benefit versus technological risk. Science **165**, 1232–1238 (1969)
39. Starr, C.: Benefit-cost relationships in socio-technical systems. In: Environmental Aspects of Nuclear Power Stations (1971)
40. Webb, G.A.M., McLean, A.S.: Insignificant Levels of Dose: A Practical Suggestion for Decision Making. National Radiological Protection Board (1977)
41. Fritzsche, A.F.: Wie sicher leben wir? Risikobeurteilung und -bewältigung in unserer Gesellschaft. Verlag TÜV Rheinland (1986)
42. Slovic, P.: Perception of risk. Science **236**, 280–285 (1987). ► https://doi.org/10.1126/science.3563507
43. Junietz, P., Steininger, U., Winner, H.: Macroscopic safety requirements for highly automated driving. Transp. Res. Rec. **2673**, 1–10 (2019). ► https://doi.org/10.1177/0361198119827910
44. Robinson, R.M., Anderson, K.J., Meiers, S.J., Browning, R.W., Alam, T.: Risk and Reliability: An Introductory Text. Risk & Reliability Associates Pty Limited (1997)
45. DESTATIS Statistisches Bundesamt: Sterbetafeln und Lebenserwartung [WWW Document] (2017). ► https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/Methoden/sterbetafeln-lebenserwartung.html. Accessed 05 Dec 2021

46. International Organization for Standardization: ISO 26262 Road vehicles—Functional Safety—Part 4: Product Development at the System Level (2018)
47. Steininger, U., Schöner, H.-P., Schiementz, M., Mazzega, J.: Validation of Assisted and Automated Driving Systems. Crash. Tech., München (2016)
48. Baum, H., Kranz, T., Westerkamp, U.: Volkswirtschaftliche Kosten durch Straßenverkehrsunfälle in Deutschland (2010)
49. Hydén, C.: The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique. Bulletin Lund Institute of Technology, Department (1987)
50. Schöner, H.-P.: Challenges and approaches for testing of highly automated vehicles. In: Langheim, J. (ed.) Energy Consumption and Autonomous Driving, Lecture Notes in Mobility, pp. 101–109. Springer International Publishing, Cham (2016). ► https://doi.org/10.1007/978-3-319-19818-7_11
51. Steininger, U., Wech, L.: Wie sicher ist sicher genug? Sicherheit und Risiko zwischen Wunsch und Wirklichkeit. VDI-Berichte (2013)
52. Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., Maurer, M.: Defining and substantiating the terms scene, situation, and scenario for automated driving. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. Presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC 2015), pp. 982–988. IEEE, Gran Canaria, Spain (2015). ► https://doi.org/10.1109/ITSC.2015.164
53. Shalev-Shwartz, S., Shammah, S., Shashua, A.: On a Formal Model of Safe and Scalable Self-driving Cars (2018). arXiv preprint ► arXiv:1708.06374
54. Nistér, D., Lee, H.-L., Ng, J., Wang, Y.: The safety force field. NVIDIA White Paper (2019)
55. Althoff, M.: Reachability Analysis and Its Application to the Safety Assessment of Autonomous Cars. Technische Universität München, Germany, Munich (2010)
56. Ulbrich, S.M.: Towards Tactical Lane Change Behavior Planning for Automated Vehicles. TU Braunschweig, Braunschweig (2018). ► https://doi.org/10.24355/dbbs.084-201809171100-0
57. Kluckhohn, C.: In: Parsons, T., Shils, E.A. (eds.) Toward a General Theory of Action, pp. 388–433. Harvard University Press (2013). ► https://doi.org/10.4159/harvard.9780674863507.c8
58. Albert, E.: Zur Wertekonzeption in den Sozialwissenschaften—Potenzial, fachliche Prägungen und mögliche Blockaden eines strategisch benutzten Begriffes (Online Publication). Sociology in Switzerland: Contributions to General Sociological Theory, Zurich (2008)
59. Thornton, S.M.: Autonomous Vehicle Motion Planning with Ethical Considerations. Stanford University, Stanford, CA (2018)
60. Graubohm, R., Schräder, T., Maurer, M.: Value sensitive design in the development of driverless vehicles: a case study on an autonomous family vehicle. In: Proc. Des. Soc.: Des. Conf., vol. 1, pp. 907–916 (2020). ► https://doi.org/10.1017/dsd.2020.140
61. Friedman, B.: Human Values and the Design of Computer Technology. Cambridge University Press (1997)
62. Christian Gerdes, J., Thornton, S.M., Millar, J.: Designing automated vehicles around human values. In: Meyer, G., Beiker, S. (Trans.) Road Vehicle Automation 6, pp. 39–48. Springer International Publishing, Cham (2019)
63. Millar, J., Paz, D., Thornton, S.M., Parisi, C., Gerdes, J.C.: A framework for addressing ethical considerations in the engineering of automated vehicles (and other technologies). In: Proc. Des. Soc.: Des. Conf., vol. 1, pp. 1485–1494 (2020). ► https://doi.org/10.1017/dsd.2020.78
64. Salem, N.F., Haber, V., Beck, H.N., Rauschenbach, M., Nolte, M., Reich, J., Stolte, T., Maurer, M.: Semantische Normverhaltensanalyse—ein Beitrag zur durchgängigen Formalisierung von Verhaltensnormen für das automatisierte Fahren (2021). ► https://www.ifr.ing.tu-bs.de/static/files/forschung/papers/download_pdf.php?id=1174-11
65. Beck, H.N., Salem, N.F., Haber, V., Rauschenbach, M., Reich, J.: Phänomen-Signal-Modell: Formalismus, Graph und Anwendung (2021). arXiv preprint ► arXiv:2108.00252 [physics]
66. Nolte, M., Rose, M., Stolte, T., Maurer, M.: Model predictive control based trajectory generation for autonomous vehicles—an architectural approach. In: 2017 IEEE Intelligent Vehicles Symposium (IV). Presented at the 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 798–805. IEEE, Los Angeles, CA, USA (2017). ► https://doi.org/10.1109/IVS.2017.7995814
67. Tingvall, C., Haworth, N.: Vision Zero-An ethical approach to safety and mobility. In: 6th ITE International Conference Road Safety & Traffic Enforcement: Beyond 2000 (1999)
68. CCA (Council of Canadian Academies): Choosing Canada's Automotive Future, Ottawa (ON). The Expert Panel on Connected and Autonomous Vehicles and Shared Mobility, Council of Canadian Academies (2021)
69. Automated Vehicles for Safety: NHTSA (n.d.). ► https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety. Accessed 03 Nov 2021
70. Owens, J.: Automated Vehicles Symposium. NHTSA (2020). ► https://www.nhtsa.gov/speeches-presentations/automated-vehicles-symposium. Accessed 04 Nov 2021
71. de Gelder, E., Elrofai, H., Saberi, A.K., Paardekooper, J.-P., Op den Camp, O., de Schutter, B.: Risk quantification for automated driving systems in real-world driving scenarios. IEEE Access **9**, 168953–168970 (2021). ► https://doi.org/10.1109/ACCESS.2021.3136585
72. Tesla: Tesla Vehicle Safety Report (2021). ► https://www.tesla.com/VehicleSafetyReport. Accessed 03 Nov 2021
73. Waymo: Safety Report and Whitepapers. Waymo (2021). ► https://waymo.com/safety/. Accessed 07 Nov 2021
74. Kronberg, P.: So, Will Autonomous Vehicles Improve Safety? (2016). ► https://www.volvogroup.com/en/news-and-media/columns/columns/traffic-safety/will-autonomous-vehicles-improve-safety.html. Accessed 04 Nov 2021
75. MADD: MADD Canada Supports Development of Automated Vehicles to Improve Road Safety (2016)
76. EndDD (End Distracted Driving): Are We Ready for Autonomous Car Features? | End Distracted Driving. EndDD (2019, October, 02). ► https://www.enddd.org/end-distracted-driving/are-we-ready-for-autonomous-car-features/. Accessed 07 Nov 2021
77. Madia, E.: Protecting Pedestrians from Themselves in the Age of Distraction | End Distracted Driving. EndDD (2018, June 10). ► https://www.enddd.org/end-distracted-driving/distracted-driving-updates/protecting-pedestrians-from-themselves-in-the-age-of-distraction/. Accessed 07 Nov 2021
78. NSC (National Safety Council): Automated & Self-Driving Vehicle Revolution: What is the Role of Government?—National Safety Council (2016). ► https://www.nsc.org/company/speeches-testimony/testimony-the-automated-self-driving-vehicle-revolution. Accessed 07 Nov 2021
79. LTAD (Let's Talk about Autonomous Driving): Partners and Community | Let's Talk Autonomous Driving. ► https://ltad.com/partners.html. Accessed 07 Nov 2021
80. LTAD (Let's Talk about Autonomous Driving): Can Fully Autonomous Vehicles Help Drive Away Stress? | Let's Talk Autonomous Driving. ► https://ltad.com/news/can-fully-autonomous-vehicles-help-drive-away-stress.html. Accessed 07 Nov 2021
81. Stockton, N.: "So, Self-Driving Cars Could Make Humans Unhealthier Than Ever. Wired. ► https://www.wired.com/story/self-driving-cars-obesity-pollution-public-health/. Accessed 05 Nov 2021
82. Groleau, C.: Engineering Researchers at UW Look at Safety and Security of Automated Vehicles | CBC News. CBC (2021, October, 03). ► https://www.cbc.ca/news/canada/kitchener-waterloo/university-of-waterloo-automates-vehicles-magna-international-1.6196110. Accessed 05 Nov 2021

83. Favaro, F.: Exploring the Relationship Between "Positive Risk Balance" and "Absence of Unreasonable Risk" (2021). arXiv preprint ► arXiv:2110.10566
84. Koopman, P., Wagner, M.: Toward a Framework for Highly Automated Vehicle Safety Validation (SAE Technical Paper No. 2018-01-1071). SAE International, Warrendale, PA (2018). ► https://doi.org/10.4271/2018-01-1071
85. Fraade-Blanar, L., Blumenthal, M.S., Anderson, J.M., Kalra, N.: Measuring Automated Vehicle Safety: Forging a Framework (2018)
86. Stolte, T., Ackermann, S., Graubohm, R., Jatzkowski, I., Klamann, B., Winner, H., Maurer, M.: A taxonomy to unify fault tolerance regimes for automotive systems: defining fail-operational, fail-degraded, and fail-safe. IEEE Trans. Intell. Veh. 1–1 (2021). ► https://doi.org/10.1109/TIV.2021.3129933
87. Homann, K.: Wirtschaft und gesellschaftliche Akzeptanz: Fahrerassistenzsysteme auf dem Prüfstand (2002)
88. Homann, K.: Wirtschaft und gesellschaftliche Akzeptanz: Fahrerassistenzsysteme auf dem Prüfstand. In: Maurer, M., Stiller, C. (eds.) Fahrerassistenzsysteme mit maschineller Wahrnehmung, pp. 239–244. Springer-Verlag, Berlin/Heidelberg (2005). ► https://doi.org/10.1007/3-540-27137-6_11
89. Wilkes, W.: Mercedes Beats Tesla to Hands-Free Driving on the Autobahn. bloomberg.com (2021). ► https://www.bloomberg.com/news/articles/2021-12-09/mercedes-beats-tesla-to-hands-free-driving-on-german-autobahn. Accessed 13 Dec 2021
90. Burton, S., Habli, I., Lawton, T., McDermid, J., Morgan, P., Porter, Z.: Mind the gaps: assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. Artif. Intell. **279**, 103201 (2020). ► https://doi.org/10.1016/j.artint.2019.103201
91. High-Level Expert Group on Artificial Intelligence set up by the European Commission: Ethics Guidelines for Trustworthy AI (2019). ► https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed 13 Dec 21

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (► http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

