The Release of Autonomous Vehicles

21

Walther Wachenfeld and Hermann Winner

21.1 Introduction

In the future, the functions of autonomous driving could fundamentally change all road traffic; to do so, it would have to be implemented on a large scale, in series production. In general, a technical system such as a car needs to be released for it to make the transition from the development phase to mass production [1]. According to the principles of project management, production release is only granted when the previously defined requirements have been fulfilled by this technical system. These requirements come from a wide range of sources, such as customers, standards or legislation. Various areas are addressed by the requirements: these include the requirements for the safety of the technical system for type approval[1] and product liability[2] reasons.

The safety of people in public road traffic is one of the oft-quoted motivations for vehicle automation, because the vast majority of present-day accidents are caused by human drivers. Based on this motivation is the requirement that substituting humans does

[1]According to Directive 2007/46/EG [2], the expression "(…) 'type approval' describes the procedure whereby one Member State certifies that a type of a vehicle (…) satisfies the relevant administrative provisions and technical requirements".

[2]Reuter [3] states: "(Tortious) product liability serves to protect any person (product users as well as uninvolved third parties) from unsafe products. Product liability regulates the compensation of damage to health or property that has been caused by a product defect".

W. Wachenfeld (✉) · H. Winner
Institute of Automotive Engineering – FZD, Technische Universität Darmstadt, 64287 Darmstadt, Germany
e-mail: wachenfeld@fzd.tu-darmstadt.de

H. Winner
e-mail: winner@fzd.tu-darmstadt.de

M. Maurer et al. (eds.), *Autonomous Driving*,
DOI 10.1007/978-3-662-48847-8_21

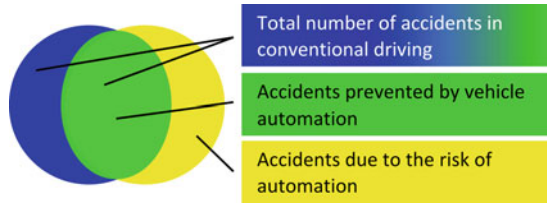

Fig. 21.1 Theoretical potential for avoiding accidents with vehicle automation [4]. Image rights: Gasser

not reduce the safety of public road traffic. This should apply to both passengers and the entire traffic system in which the autonomous vehicle is in motion. What this requirement means, and whether it would actually be fulfilled upon the introduction of the autonomous vehicle, is the focus of the following discussion.

A starting point is provided by Gasser et al. [4]: The report from the German Federal Highway Research Institute considers the development of accident numbers upon the introduction of vehicle automation. Starting from the total number of accidents for conventional driving (see Fig. 21.1, blue and green field) it is assumed that accidents (green field) are avoided by means of vehicle automation. However, new accidents could also be caused by the risks of automation (yellow field).

This representation does not differentiate in terms of the severity of the accident, but the severity of the accident is also relevant when considering the impact on safety. Safety is generally described as the absence of unreasonable risks. This risk is defined as a product of the probability of an accident and the severity of that accident.

Figure 21.2 illustrates in a qualitative way this theoretical risk avoidance potential depending on the severity of the accident. Here Fig. 21.2 adheres to the findings of Heinrich [5] and Hydén [6] that accidents of decreasing severity occur in larger numbers. The scale of the related severity of the accident is ordinal, meaning that there is clearly an order between the different degrees of severity: For example, a fatality is weighted as graver than a serious injury. However, academics are divided on the relative weighting of these different degrees. While degrees of severity are compared in terms of costs, this is contentious and will not be discussed further in this work.

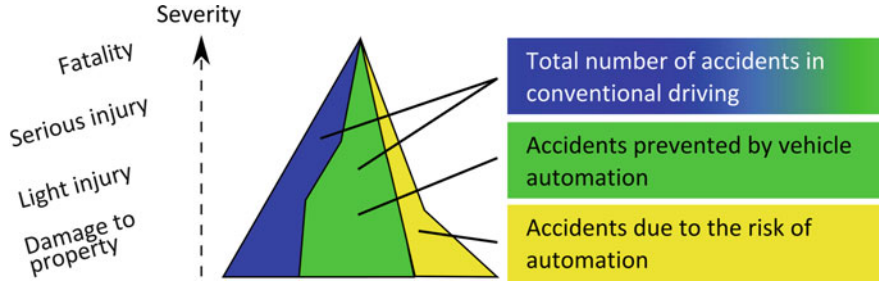

Fig. 21.2 Theoretical potential for avoiding accidents with vehicle automation with consideration of gravity of accident (similar to [4]). Image rights: Author has copyright

Considering the severity and the number of accidents shows that while risks are removed (Fig. 21.2 green area), there are risks remaining (Fig. 21.2 blue area) which are not addressed by vehicle automation. In addition, new risks are created by the substitution of humans and the automated execution of the driving. The human is no longer available as a backup in the case of a failure or a defect. The yellow area in Fig. 21.2 illustrates this additional risk. It is uncertain here whether the removal of risks and the creation of additional risks is uniform across the degrees of severity. It is possible that there is a greater reduction in serious accidents but an increase in less serious accidents. Figure 21.2 illustrates this idea via the deformation of the assumed triangle.

For the approval of fully-automated driving, this means that not only a reduction in the number of accidents must be proven, but rather an accepted ratio V_{acc} between avoided R_{avo} and additionally caused risks R_{add}.

$$V_{\text{acc}} = \frac{R_{\text{add}}}{R_{\text{avo}}}$$

Contrary to many assertions, it has not yet been proven that a ratio of less than 1 is actually necessary for the approval. Will autonomous vehicles actually increase traffic safety? If the ratio were greater than 1, the system would reduce traffic safety. Examples exist today whereby corresponding added benefits create acceptance for additional risks: For example, for many motorcyclists the experience of freedom, driving pleasure, etc., balances out the considerable additional risk compared to other means of transport. In addition the apportionment of benefits and risks facilitates the acceptance of motorcycling. The added benefits and the additional risk mainly affect the person on the motorcycle. The risk for other road users created by a motorcycle lies between the risks created by a bicycle and a car, and therefore motorcycling is acceptable without added benefits for other road users.

In this document, *no* specific value is determined for the acceptable ratio for autonomous driving, because this value is the result of a complex discussion among those who would be affected by autonomous driving. This value varies depending on various factors such as societal, political and economic differences. A vivid example of this is the acceptance of the use of nuclear energy in Germany, the USA or Japan in the last years: On the one hand, the accepted ratio varies considerably between the countries, and on the other, this changes over time, so that for example in Germany in 2012, a nuclear phase-out was decided on.

The central component of this document is the evaluation of autonomous driving, i.e. the study of the methods that are to enable safety assessment of autonomous vehicles. Even though a large number of papers describe the potential of autonomous driving in theory, the authors are not aware of any document that has conducted this evaluation. In order to show why this is so, we will first describe the current release concepts in the automobile industry, and then show what the requirements are for test concepts. In the

third section, we describe the special features of autonomous driving in relation to current systems. On the basis of this, the fourth section looks at the special challenge for the production release of autonomous vehicles. The approaches that address this challenge are discussed, and then a conclusion on the production release of autonomous vehicles is drawn in the final section.

21.2 Current Test Concepts in the Automobile Industry

The safety validation concepts currently used in the automobile industry are for obtaining approval for four distinct automation levels. To illustrate the difference for the test of these systems compared to autonomous driving, these four systems will be explained briefly.

The first system in series is the driver-only vehicle without the automation of the driving task. For these systems, it can be seen that, on the one hand, the components used do not exceed maximum failure rates, and on the other, that the driver is able to maneuver the vehicle reliably in road traffic (controllability). Here the abilities of the driver are relied on, as the results of the conducted tests with test drivers are transferred to future users in the subsequent area of use. Over the last decades, this has shown itself to be successful in serving as proof of safety. Despite the increasing number of kilometers driven in road traffic, the number of accidents remains constant, and the number of fatalities has even fallen.

The second level of automation in series is the assisting system: For systems such as Adaptive Cruise Control (ACC) or Lane Keeping Assist (LKA), their functions have to be covered by the test in addition to the existing scope of testing. The option of a take-over by the driver and controllability must be provided in systems that actively support the driving task, increase comfort, and reduce the burden on the driver. The Code of Practice [7] thus assumes that, in this Advanced Driver Assistance System (ADAS), responsibility for vehicle behavior remains with the human driver. For these systems it also applies that the abilities of the driver are relied on, so that the results of the conducted tests with test drivers are transferred to future users in the subsequent area of use.

The first partly-automated systems have also been approved for use in series cars: Depending on the speed, ACC in combination with LKA takes over the lateral and longitudinal control for the driver. According to the definition, in the third category of systems, the driver is also responsible for the vehicle behavior. Therefore, this test also focuses on the possibility for the take-over and the controllability by the driver; and so the same principle applies as with the assisting system, which relies on the abilities of the vehicle driver to correct undesired automation behavior. This level of automation presents the special challenge for the safety validation that results from the conflict between relieving the driver and the necessary situation awareness of the supervisor of the lateral and longitudinal control. However, here too the driver is ultimately responsible.

Of particular interest for the test are emergency intervening systems, which automatically intervene in the vehicle control and thus in the vehicle dynamics. The goal of this fourth category of systems is to counter the driver's loss of control over the situation. For

example, Electronic Stability Control (ESC) and Emergency Brake Assist (EBA) are components of mechatronic brake systems that apply additional or reduced braking force without any action on the part of the driver, thus actively intervening in the vehicle dynamics. This is performed during the driver's loss of control, when the vehicle, in combination with the driver, is at a higher level of risk. ESC is designed in such a way that an intervention is carried out when the driver clearly no longer has control over the vehicle in the current situation (e.g. in the case of extreme over- or understeering). In contrast, the EBA becomes active when the reaction time and the braking distance before a rear-end collision are no longer sufficient for a human to prevent this accident. The goal of validating the system regarding safety requirements is to show that emergency intervening systems should only become active (true-positive) when the loss of control becomes obvious and thus there is a severely increased risk. For this, it must be shown that the false-positive rate becomes as small as possible and/or the effects can be controlled by the driver; the false-positive and false-negative rates of the EBA mainly depend on the object detection. Figure 21.3 shows a Receiver Operating Characteristic curve (ROC curve), which describes this relationship for the object detection.

As these emergency intervening systems are systems with no guaranteed operation, an increase in safety can be achieved by reduced usage combined with a smaller false-positive rate. Additionally, these systems enable overriding. ESC and EBA employ the selective braking of wheels to intervene mainly in the braking system, and various strategies can be used to override them, by steering and/or accelerating.

As has been shown, the main focus in the development of the four system levels is controllability by the driver. The goal is either to enable controllability for the driver or to restore it for him/her (design for controllability). Therefore, the driver as a backup is the basis for validating current vehicles regarding safety and hence also for the production release.

The development and verification of this controllability for the driver is generally carried out in accordance with the procedure model in Fig. 21.4. This procedure based on the V-Model differentiates between the downward branch on the left—development and

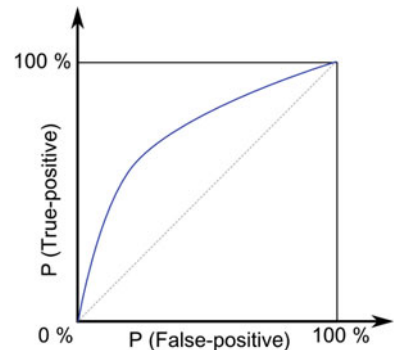

Fig. 21.3 Representation of the principle of a receiver operating characteristic curve based on Spanfelner et al. [8]. Image rights: Author has copyright

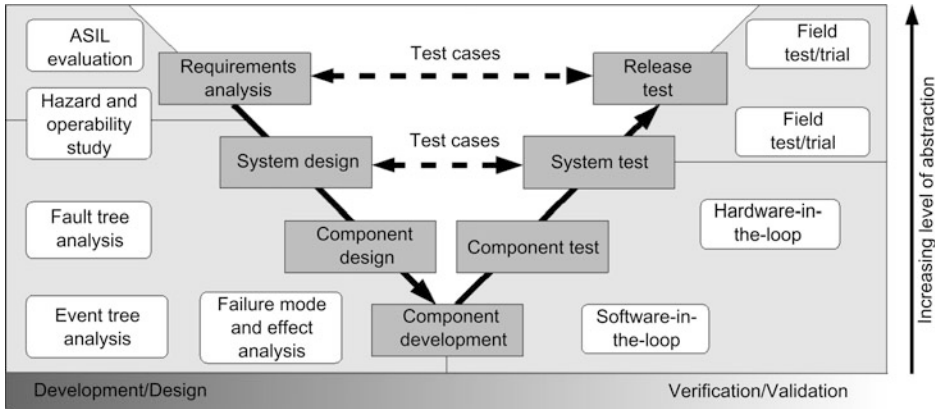

Fig. 21.4 Safety evaluation methods in the development process (according to [9]). Image rights: Alexander Weitzel et al. Federal Highway Research Institute report: Absicherungsstrategien für Fahrerassistenzsysteme mit Umfeldwahrnehmung

design—and the upward branch on the right—verification and validation—as a means of quality assurance. A test concept is followed for the quality assurance.

As shown by Schuldt et al. [10] in Fig. 21.5, a test concept comprises the analysis of the test object (object under test—OUT), the test case generation, the test execution and the test evaluation.

The analysis of the test object and the test case generation should be performed during the development/design phase, so that the test cases to be carried out are already defined for the verification and validation (see Fig. 21.4 procedure model). According to Horstmann [11] and Weitzel [9], at present a distinction is made between three methods for the determination of test cases: One method is the test specification based on the specification sheet, whereby test cases are defined based on system specifications, which have been set down in specification sheets. The second method is the risk-based test specification, whereby risk considerations are used to determine the test cases. The third method is the interface-based test specification, whereby the test cases are selected in

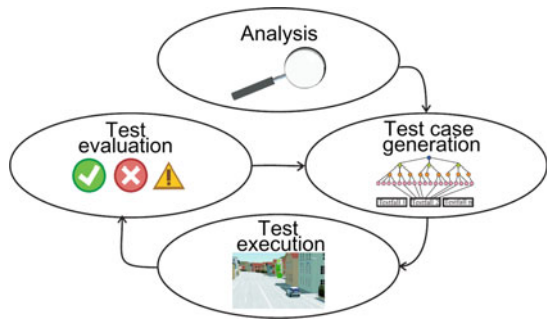

Fig. 21.5 Procedure for test concept (according to [10]). Image rights: Schuldt Braunschweig

order to cover the value ranges of the interfaces. For all these methods, the driver-vehicle system is the basis of the test case determination.

To start with the quality assurance as early as possible, tests are already carried out in virtual test environments before the first test vehicles are ready for testing. The test execution by means of model- and software-in-the-loop tests works based on simulation models of the vehicle, the human and the environment. The test cases previously identified are used here. The further the development progresses, the greater the number of real components available for testing. Test benches, driving simulators or testing grounds are used in these tests. The tests performed using hardware-in-the-loop, driver-in-the-loop or vehicle-in-the-loop provide information about the quality of the components and functions being tested. To check the actions and reactions of the driver-vehicle-environment system (to close the loop), simulation models are also needed in performing these tests. Therefore, simulation models will be required continuously for the test execution up to this development point in order to test the entire vehicle. Simulation models are mappings of reality in software and have per se the property of simplifying the real world.

As a result of this fact, there currently exists no safety-relevant function in a series vehicle that has not also been tested with real test vehicles. Thus, for testing current systems, the automotive industry always falls back on real vehicles, real humans and a real environment.

A result of the necessary use of real driving is, for example, that before the production release of the Mercedes Benz E-Class (W212), a total of 36 million test kilometers were completed [12].[3] According to Fach et al. [13], the safety validation of a current driver assistance system alone requires up to 2 million test kilometers. After 50,000 to 100,000 km were covered in these test drives between two interventions of the first level of the EBA, this high number of test kilometers becomes understandable. This does not even consider the fact that the more critical second level of the EBA was not triggered during these test kilometers (compare assertion in Fig. 21.3). This eight-figure total of test kilometers is accompanied by considerable costs for the vehicle prototypes, test drivers, test execution and the evaluation of same. While the time requirement can be reduced by means of parallel testing with multiple vehicles, additional costs are incurred for the vehicle prototypes.

This example shows that even for current driver assistance systems, validating safety based on real driving in road traffic represents an economic challenge for the OEM (Original Equipment Manufacturer). This challenge grows further against the background of the increasing number of functions and widening ranges of variants and versions for each vehicle model. For example, Burgdorf [14], deduces a number of $160 \cdot 2^{70}$ variants for the BMW 318i (E90) with components such as body form, engine, transmission, drive, color, A/C, infotainment.

[3]"(…) The [E-Class] arrived by way of comprehensive virtual tests with digital prototypes and a total of 36 million test kilometers (…)." (Retrieved 28/07/2014).

Therefore, there are already endeavors to use other test execution tools alongside real driving for final safety validation. The only example of this known to the authors is the homologation of ESC systems. According to ECE Regulation 13H for the EU [15], there is the option to perform some of these tests in the simulation:

> When a vehicle has been tested physically in accordance with section 4, the compliance of other versions or variants of the same vehicle type can be proven by means of computer simulations that adhere to the test conditions of section 4 and the test procedures of section 5.9.

Note that this only applies to the ESC system. As an example, Baake et al. [16] describe the homologation of ESC systems for vans from Mercedes-Benz in collaboration with Bosch and IPG CarMaker: Using what are known as master cars, a vehicle model was created in CarMaker, and these master cars were used to collect reference data on the basis of which the simulation model was validated. This enabled the simulation-based recommendation for the approval of further vehicle variants with different settings. Baake et al. also report on the transfer of this procedure to the Cross Wind Assist (CWA) function, although this has not yet been done.

21.3 Requirements for a Test Concept

In order to discuss in the following section why full automation poses a particular challenge for safety validation, we will first describe the requirements for test concepts to assess safety. These are divided into effectiveness and efficiency criteria.

21.3.1 Effectiveness Criteria

Representative—valid
The requirement for representativeness has two aspects: On the one hand, the test case generation must ensure that the test coverage required is achieved. For example, a vehicle should not only be tested at 20 °C and sunshine, as it will be exposed to snow, rain and temperatures under 0 °C in real situations. Additionally, vehicle limit samples (tolerances during production) should be considered in the test case generation. On the other hand, the test execution must encompass the minimum degree of reality required. This means that the simplification in the representation of reality must not influence the behavior of the OUT nor the behavior and properties of the environment with respect to real behavior.

Variable
The test execution must provide the option to implement all the test cases defined by the test case generation.

Observable
For the test evaluation in particular, it is necessary to observe parameters of the test execution. Only when the situation can also be described it is possible to make the statement test "passed" or "not passed".

21.3.2 Efficiency Criteria

Economical
There are two parts to the requirement for the economical test concept: On the one hand, the test execution should be prepared and carried out as quickly as possible in order to be able to provide the persons involved in the development with feedback on the test object immediately. On the other hand, it must be ensured that the test execution is prepared and carried out at the lowest cost possible.

Reproducible
Reproducibility greatly reduces the work required for regression tests. For example, if an error has been detected and the test object modified accordingly, the goal is to subject the OUT to a test in the same scenario as before.

In good time
The earlier in the development process that a product can be tested informatively, the fewer the development steps that need to be repeated in the case of an error.

Safe
The test execution should not exceed the accepted risk for all participants. This must be considered in particular for real driving, whereby road users are participating in the test without their knowledge.

The requirements described are fulfilled sufficiently by the current test concepts, and therefore the four different automation levels presented are approved. However, the recalls of all the OEMs, which affect millions of vehicles, indicate that these test concepts certainly do not address everything. Are these concepts also suitable for validating the safety of new systems such as autonomous driving for public road traffic? Nothing changes about the requirements presented. However, as will be described in the following section, the OUT changes greatly.

21.4 Special Features of Autonomous Driving

In the following section, the difference between fully-automated driving and current driving in road traffic is explained. After this, the differences between the traffic systems for air travel, rail travel and road traffic are present in compact form, so that only limited findings from these areas can be transferred.

21.4.1 Comparison Between Current Automation and Full Automation of Road Vehicles

For the previously described safety validation of the four levels of automation available in series, the focus was on the vehicle, and in particular on its controllability by the driver. In the combined representation of the three-level model for human target-oriented behavior based on Rasmussen [17] and the three-level hierarchy of the driving task based on Donges [18] in Fig. 21.6, this validation corresponds to the elements with the green background. The vehicle and its behavior in the longitudinal and lateral directions are tested; in the process, the behavior or abilities of the future driver are not tested, but only the possibilities for the test driver to control the vehicle in the test cases by means of steering and acceleration control. Therefore, the green box only overlaps slightly with the area that stands for the driver.

For full automation, the abilities of the driver are now omitted and he/she also no longer functions as a backup. The driving task, i.e. navigation, guidance and stabilization/control, is taken over by the driving robot. This means that for autonomous

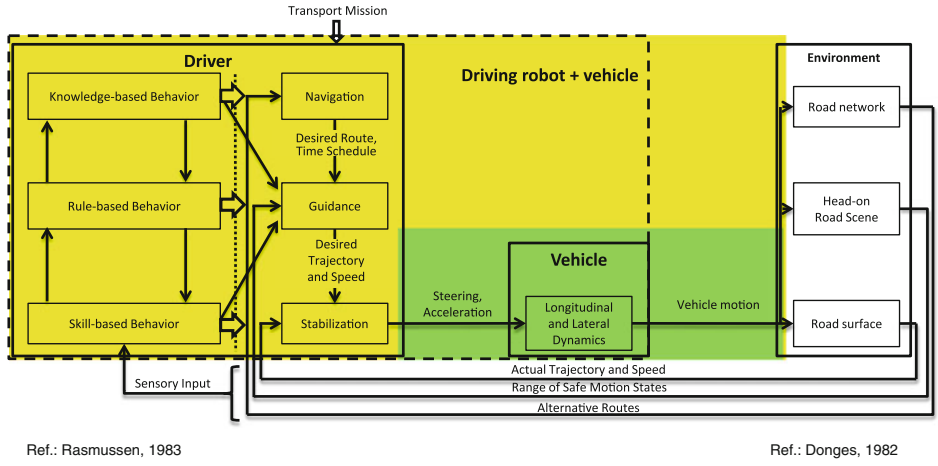

Fig. 21.6 Three-level model for human target-oriented behavior based on Rasmussen and the three-level hierarchy of the driving task based on Donges [18]. Image rights: Donges?—to be clarified

driving, there is no test of the controllability, but only a test of the operation of a technical system. On the one hand, this makes the test easier, because the uncertainties due to the human and its individual differences no longer need to be covered by the test. On the other hand, there is no longer the option to use test cases and test drivers to draw conclusions about other use cases. The human, who generally acts based on skills, rules and knowledge, is omitted.

For the safety validation of current systems, safety must be proven that results from the driver and the vehicle in combination; however, for the production release of the vehicle, at present the focus is solely on the vehicle. Additionally assumed, but not tested, is the "reliability" of the driver. In assessing the autonomous system in terms of safety, the safety now results exclusively from the technical system of the driving robot and the vehicle (yellow field of Fig. 21.6), which must be proven.

Figure 21.6 shows on the one hand that here the quantity of tasks that must be tested increases: The driving robot is required for a wide variety of application areas (see Use Cases Chap. 2) such as navigation, guidance, stabilization/control. This task quantity presents a particular challenge in public spaces without access limitations. On the other hand, the task quality of the technical system changes. Current systems are merely executive or are continuously monitored by a human, while for the autonomous system the execution of a task must fulfill the requirements of the safety discussed at the beginning of this document.

21.4.2 Comparison of the Stipulations in Air Travel, Road Traffic and Rail Travel

Along with road traffic, there are other traffic systems in which automation has established itself. However, the following section will discuss the extent to which the challenges and solutions from these areas are transferable to vehicle automation.

The automation in (civilian) air travel does not currently provide any examples of full automation. Even if pilots only very rarely actually perform flying tasks, they are still present in a supervising and operating capacity. Table 21.1 provides an overview of the differences in the traffic systems, which was taken from Weitzel et al. [9] and Ständer [19]. For the safety validation, the safety concept for the traffic flow is of particular interest here, as this shows the differences between air travel and road traffic. Air travel operates in a legally self-contained traffic space, a collision warning system is mandatory, and external monitoring of operations is provided by air traffic control.

The railway traffic system provides examples of full automation: For example, an automated underground railway is in operation in Nuremberg. However, according to Table 21.1, even in this traffic system the safety concept for the traffic flow in particular differentiates between road traffic and the railway. There is a legally self-contained traffic space for rail travel; in addition, logic-based systems and external monitoring are used to avoid a collision between two trains.

Table 21.1 Comparison of the conditions in the traffic systems, taken from Weitzel et al. [9] and based on Ständer [19]

	Air travel	Road traffic	Rail travel
Movement options	3-D (space)	2-D (area)	1-D (line)
Operator			
Responsible vehicle operator	Usually redundant	Not redundant	Not redundant
Professionalism of the vehicle operator	Almost completely full-time occupation	Small proportion full-time occupation	Almost completely full-time occupation
Training			
Theory	> 750 h	> 21 h	∼ 800 h
Practice	> 1500 h	> 9 h	∼ 400 h
Training for vehicle type	Yes	No	Yes
Further training	Required	Not required	Required
Safety concepts of the traffic flow			
Traffic space self-contained	Legally defined boundaries	In special cases	Legally defined boundaries
Driving by sight	No, only in special cases	Yes	No, only in special cases
Technical equipment (examples)	Collision warning systems mandatory	Road markings, traffic lights, traffic signs	Automatic vigilance device, intermittent train control, automatic driving and braking controls
External monitoring	Yes, air traffic control	No	Yes, centralized traffic control, operation center
Technical framework			
Documentation of tours/operating hours	Yes	No	Monitoring of operating performance, automatic tachograph
Servicing, repairs	Only by certified companies	Workshops, DIY	Only by certified companies, and then also small workshops
Accident analysis	Every accident/serious malfunction, by independent state-run body	In individual cases, by certified assessor	Every accident/serious malfunction, by independent state-run body
Number of vehicles (in Europe)	10^3 (decreasing)	10^6 (increasing)	10^3 (decreasing, with increasing kilometric performance of each traction unit)
Change of model	Approx. 20 years	Approx. 5–7 years	Approx. 20 years for traction units

As a mixed operation, road traffic does not fulfill the condition of a self-contained traffic space and external monitoring. The differences show why solutions for the production release cannot be transferred directly to autonomous driving.

This comparison should not exclude the possibility that all solutions from air travel and rail travel are of no interest for road traffic. Certainly similar problems exist, such as the reliability of safety-relevant components.

21.5 The Challenge of Releasing Fully-Automated Vehicles for Production (the "Approval-Trap")

As has been shown, the functions of autonomous driving as an OUT differ fundamentally from current road vehicles, but also from means of transportation in air and rail travel. Therefore, we now want to determine how meaningful the current test concepts presented would be when transferred onto autonomous driving. We will also discuss what the effects would be of continuing with the current test concept.

21.5.1 Validity of the Current Test Concept for Autonomous Driving

It has already been explained that a test concept consists of test case generation and test execution. Now we want to discuss how and whether both are transferable to autonomous driving.

Test case generation
The three procedures for test case generation have already been explained briefly in Sect. 21.2; these procedures are based on the assumption of the driver's driving capability. The question of whether a random driver can control the test object is tied to the legally stipulated driver's license. According to the Road Traffic Act (§ 2 Abs. 2 StVG), this driver's license is only issued if, among other things:

- the applicant has attained a minimum age,
- he/she is suitable for driving a motor vehicle,
- he/she has received training,
- and has passed theoretical and practical tests.

And according to § 2 Abs. 4 StVG, suitable is taken to mean:

> A person is suitable for driving motor vehicles if he/she fulfills the necessary physical and mental requirements and has not substantially or repeatedly contravened traffic regulations or criminal laws.

On the basis of this required driving capability on the part of the driver, the test case generation is limited to example situations: It is assumed that when the test driver has

mastered these example situations, he/she and every other driver with a driver's license will also master the other relevant non-tested situations when driving. These include situations in which the driver is actively driving, but also those situations in which the driver is supervising the system and, if necessary, takes over control. Therefore, in combination with the driver's license test, these test cases provide a metric that allows a conclusion to be drawn about the safety of the driver-vehicle system. The way in which it would be possible to optimize the practical driver's license test as an evaluation basis for assessing the driving capability is discussed by Bahr [20].

In the absence of the driver, the currently accepted metric no longer applies, and therefore the reduction of the test cases is no longer admissible. The test case generation for autonomous driving must cover the driving capabilities in particular—a new quality of functions—which the human previously brought to the driver-vehicle system. The theoretical and practical tests of the driver's license test do not represent the difficulty here. However, the following paragraphs—§ 10 Minimum Age, § 11 Suitability and § 12 Visual Faculty of the Driver's License Regulation—present the challenge. Therefore, these paragraphs stand implicitly for comprehensive requirements for the properties of the humans who perform driving tasks. The human who fulfills these requirements has

- experienced hundreds of thousands of kilometers as a road user,
- experienced social behavior as a member of society,
- learned cognitive abilities,
- trained sensomotor abilities,
- etc.

The authors are not currently aware of any method for validly testing these functions for a technical system. Therefore, the accepted metric and the reduction of the test cases no longer apply if the human is removed from the responsibility of performing the driving task. The current test cases are not meaningful for releasing automated vehicles for production, and therefore the test case generation must be adapted to the new system.

Test execution

As has already been shown, different methods ranging from HiL to SiL to real driving are used for the test execution. At present, real driving is the most important method for the approval; the reason for this, in particular, is the validity combined with the justifiable economic overhead. However, along with the economic overhead, autonomous driving also presents a systematic challenge for the known methods. At present, real driving stands for driving in public road traffic with test drivers. The task of the test driver is to drive or supervise the vehicle in every situation in accordance with the task of the vehicle user. Transferred to autonomous driving, the use of a test driver in the driver's seat would be non-real behavior of a user, as the user does not have to supervise the vehicle and the environment anymore and intervene. Additionally, the vehicle could also participate in the road traffic without passengers (depending on the use case), and therefore a test driver would represent a non-real component in the vehicle. As a result, there is a risk that the

use of a test driver could influence the other road users and alter their behavior. Further reflections on this topic can be found in the Chap. 7.

Therefore, along with the test case generation, the current test execution is not directly transferable to autonomous driving.

21.5.2 Millions of Kilometers on Public Roads Until the Production Release of Fully-Automated Vehicles

The following theoretical consideration will show what it means to retain the current test concept despite the differences shown. Let us assume that a reduction in the test cases is not possible for autonomous driving, because no method exists, as with the driver's license test for humans. The objective is still to draw a conclusion as to whether the risk is increased or not by the use of the autonomous vehicle:

$$V_{\text{acc}} = \frac{R_{\text{add}}}{R_{\text{avo}}} < 1$$

Here we should note once again that this condition is in no way imperative. However, for the theoretical consideration, a condition of less than 1 is assumed to be the worst case scenario.

The only metric that the authors are aware of that can be used to determine such a relationship are the figures from the subsequent evaluation of traffic accidents. For Germany, these are the figures from the Federal Statistical Office. For example, for 2012 the Federal Statistical Office [21] cites 3375 fatal accidents recorded by the police in Germany. The figure for fatalities is used because this represents the worst case scenario for the verification required. With a total of 709 billion km driven in Germany, this figure represents an average of 210 million km between two fatal accidents. As these figures only represent an expected value, shorter or longer distances also exist between two accidents. To represent this distribution of the accident events, we use the Poisson distribution:

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Here it is assumed that the occurrence of an accident is an independent and non-exhaustive random process $P_\lambda(k)$. In the equation, k corresponds to the number of accident events and λ to the expected value with which this event occurs. The expected value λ is defined by the quotient

$$\lambda = \frac{s_{\text{test}}}{s_{\text{perf}}},$$

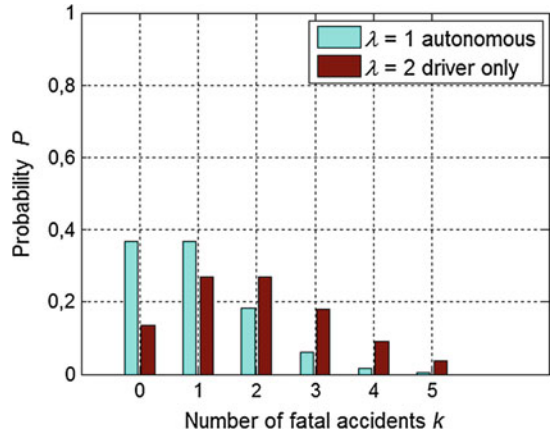

Fig. 21.7 Poisson probability distribution for the number of accidents with different expected values. Image rights: Author has copyright

whereby s_{test} stands for the observed test kilometers and s_{leist} for the performance of the system. The performance stands for the expected number of kilometers between the accidents. The probability distributions for $k = [1\,2\,3\,4\,5]$ and $\lambda = [1\,2]$ are shown in Fig. 21.7 as an example.

The figure clearly illustrates the problem of providing verification of a certain level of risk: Let us assume that the blue distribution stands for an autonomous vehicle and the red distribution for a driver-only vehicle. Both vehicles are driven the same number of test kilometers $s_{\text{test}} = a_s \cdot \bar{s}$, with the distance factor $a_s = 2$ and the average interval $\bar{s}$ between two fatal accidents. The performance $s_{\text{perf}} = a_{\text{perf}} \cdot \bar{s}$ of the autonomous vehicle is greater than that of the driver-only vehicle by the performance factor $a_{\text{perf}} = 2$. Consequently, for the autonomous vehicle $\lambda = 1$, and for the driver-only vehicle $\lambda = 2$.

Even though the autonomous vehicle is characterized by double the performance of the driver-only vehicle according to the previous assumption, during the test the autonomous vehicle was involved in a fatal accident (probability $P_1(1) = 1 \cdot e^{-1} \approx 0,37$), but not the driver-only vehicle (probability $P_2(0) = 1 \cdot e^{-2} \approx 0,14$). Therefore, a conclusion that the autonomous vehicle is less safe than the driver-only vehicle must be called into question. In any case, this example shows that a distance factor a_s greater than 2 is necessary to be able to draw a conclusion with a sufficiently high significance about the performance of autonomous driving.

From a scientific point of view, for example, an error probability of 5 % must be assumed, and therefore the same significance level $e = 5\,\%$ must be used. A correspondingly large distance factor a_s must be selected, depending on the number of accidents, in order to have a probability of less than 5 % for a vehicle with a lower performance to achieve this low number of accidents. Figure 21.8 shows the result of this consideration and the numeric calculation of the values.

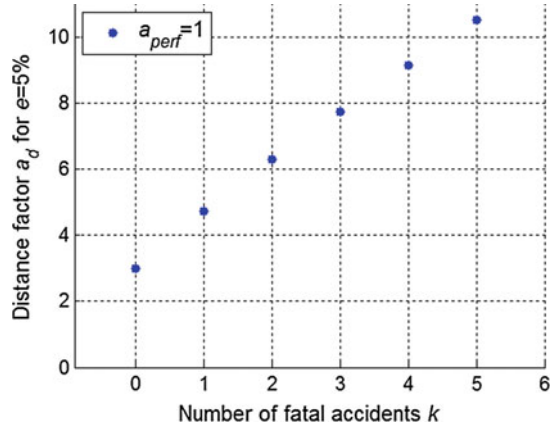

Fig. 21.8 Distance factor at significance level 5 %. Image rights: Author has copyright

The data point at zero fatal accidents means that, with a distance factor of $a_s \approx 3$, the probability is less than 5 % that a worse vehicle than the comparison group is not involved in a fatal accident.

Unfortunately, the probability of success for this test is just as small. Because if the test vehicle is just as good as the comparison group, i.e. performance factor $a_{\text{perf}} = 1$ applies, it follows that the probability of success for this verification is also only 5 %. For the test to be successful, a greater probability of success is desirable. As an example, a probability of success of 50 % is now demanded; by which a test shows that the test vehicle is not worse than the comparison group. For this, the test vehicle must perform better than the test group. Figure 21.9 shows the result of this consideration.

The first point expresses the following: If the test vehicle is approx. 4.3 times better than the comparison group, the test is successful with a probability of 50 % that the test vehicle with an error probability of 5 % is better than the comparison group.

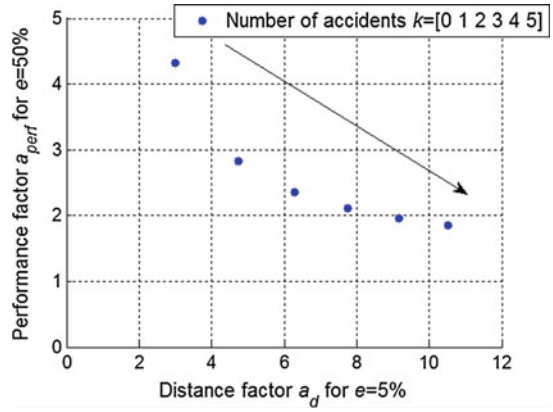

Fig. 21.9 Distance factor over performance factor at a significance level of 5 % and a probability of success for the test of 50 %. Image rights: Author has copyright

What this result now means for the test drive with the autonomous vehicle is demonstrated by the distance of 210 million km between two fatal accidents. The last point in Fig. 21.9 expresses the following: If the autonomous vehicle is approx. twice as good ($a_{\text{perf}} \approx 2$) as the comparison system (current vehicles), a test distance of at least 2.1 billion km must be driven ($s_{\text{test}} = a_s \cdot 210\,\text{Mio km}$). In this case, the verification has been achieved with 50 % probability, but five accidents would also occur with the same probability.

Ironically, it follows from this consideration that the easier the vehicle driving is, the greater the number of test kilometers that must be driven, as the comparison value is correspondingly higher. For the interstate pilot, the current figures of the Federal Statistical Office indicate a comparison value of 662 million km between two fatal accidents. Accordingly, 6.62 billion test kilometers must be driven on the interstate in order to correspond to the presented conditions.

This theoretical excursion into statistics shows that production release can become a challenge, if not an actual trap, for autonomous driving. Hereby, a number of factors for determining the test kilometers have not been addressed yet; for example, a variation of the system would mean that the test kilometers would have to be driven again, or the test with and without passengers could use a factor of two in the calculation. The effect on the determined necessary kilometers of different parameters not considered here such as area of use, accident severity, accident cause and comparison vehicle, is derived in detail in Winner [22].

These considerations are theoretical observations with freely made assumptions. However, this approach is still suitable for illustrating the problems and challenges, and for motivating the approaches that follow here.

21.6 Possible Approaches for Solving the Challenge of Testing

As has been shown, autonomous driving represents a new OUT which, due to its properties, calls the classic test concepts into question. New approaches are required to overcome the testing challenge described: Accordingly, the next section will discuss why reusing approved functions, and thus an evolutionary approach, seems necessary from the perspective of safety validation. After this, we will discuss existing approaches that could speed up testing.

21.6.1 Reusing Approved Functions

The first and simplest possibility of obtaining the production release for a new system is in reusing functions already released. If a system is used in the same way as before, a release already issued can be taken over. However, if the scope of functions is expanded, this must be treated again; the smaller the new area involved is, the less work is required.

Based on this argument, an evolution across all dimensions would seem to be a possible approach for overcoming the testing challenge. Dimensions here refers, for example, to the speed, the area of use but also the degree of automation. A distinction can be made between two perspectives in selecting the evolution steps: From the perspective of a function developer, due to the reduced speed and the limited access to the scene, the interstate during a traffic jam is a suitable starting scenario. From the perspective of the previously presented statistical considerations, a meaningful starting scenario would be one in which the human as a comparison group would perform as badly as possible, i.e. making as many errors as possible. As many errors as possible means a short distance, making the verification of the performance easier.

The revolutionary step—an autonomous vehicle without evolutionary intermediate steps—contradicts this approach and seems unlikely.

21.6.2 Speeding up the Testing

Despite the evolutionary approach, the safety of new functions still has to be validated. To speed this up, there are basically two adjustments that can be made: Firstly, the What can be changed, and secondly the How. What test cases need to be inspected, and how will these tests be performed? Schuldt et al. [10] call this the test case generation and the test execution.

21.6.2.1 Test Case Generation

The test case generation defines the tests to be carried out. According to Schuldt et al. [10], the large number of influencing factors in the area of use and their value ranges result in a conspicuous number of test cases. As already described, the systems currently in use are based on the capability of humans and their options for controlling the vehicle. This results in a stark reduction in the test cases theoretically required. Therefore, a metric exists that enables a conclusion about the safety without testing all the situations. This reduction does not apply for the autonomous vehicle, and therefore new ways must be found of reducing the number of test cases for the autonomous vehicle. During the test case generation, the requirements for a test concept detailed in Sect. 21.3 must be considered. In particular, the representativeness is at risk when test cases are omitted.

Here the approaches from Glauner [23] and Eckstein [24] describe the identification of relevant or critical situations in public road traffic. Based on previously defined event classes, potential critical situations are identified during the test drives or large-scale field studies. These critical situations are incorporated into the test case generation, and less critical situations can be omitted as a result. This reduction is based on the assumption that situations that are less critical are covered by critical situations. A task that remains unsolved at present is the search for a valid measure of risk that enables an evaluation in the first step, and the selection of critical situations in the second step.

Another procedure for reducing test cases is provided by Schuldt et al. [10]: A generic test case generation is proposed to cover factors influencing the safety ensured by the

system as sufficiently as possible. This should use black-box testing procedures and combinatorics, and also be low-redundancy and efficient. This approach is based on statistical considerations without knowledge and experience of the test object, but it still has the potential to reduce the test cases required.

The approach described by Tatar and Mauss [25] is also suitable for black-box testing: an optimization is used for the generation of test cases. Here the input variables of a XiL simulation are varied in such a way that the evaluation function to be defined for the test is optimized. Despite the challenge of the valid XiL simulation and the required evaluation function, this approach provides the option to focus the test cases on those evaluated as relevant.

A fourth theoretical approach is to use and test a safety concept using formal methods [26]. Similarly to the human assumed to be a monitor and a part of the safety concept of current vehicles, a verified reliable safety concept could make testing the overall functionality of the vehicle in its complete representativeness superfluous. This would make a reduction of the test cases possible.

21.6.2.2 Test Execution/Test Tool

Along with the possibility of reducing the test cases during the test case generation, the test execution also has potential for speeding up the process. However, if we deviate from real driving and select another testing tool for the test execution, there is always an attendant simplification. This is described in more detail by means of Fig. 21.10.

Figure 21.10 divides the testing tools into nine classes which are differentiated based on how the vehicle and the environment are represented. The passenger is assigned to the vehicle in this representation, as he/she is situated in the vehicle and does not actively intervene in the autonomous driving.

Real driving represents both the environment and the vehicle in reality. Accordingly, during these tests there is the risk of real accidents and their consequences. The environment is not controlled, and this results in test situations based on the randomness of reality; accordingly, the reproducibility for complex situations with other road users is not a given. This testing tool can be used, at the earliest, with the first roadworthy prototypes, and therefore occurs at the end of the development process.

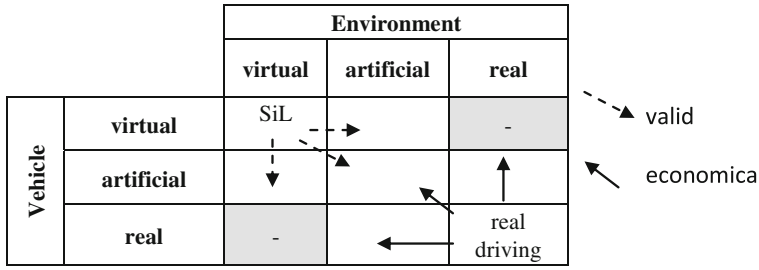

Fig. 21.10 Classification of testing tools for testing autonomous vehicles. Image rights: Author has copyright

An alternative is to test real vehicles in an artificial environment: This corresponds to driving on a test ground, as there situations are created artificially on the one hand, and on the other the "road users" are conscious of being involved in a test. Reality is simplified for the benefit of safety, variability, observability and reproducibility. From economic perspectives, while the test cases are tested specifically and do not have to be experienced randomly as in real driving, setting up the test field requires additional time and financial resources.

Additionally, an artificial vehicle could move within a real environment; in this case, artificial refers to equipping the autonomous vehicle with a supervisor, for example, that has the option to intervene in the driving task. This could be a test driver with a steering wheel and pedals, or alternatively a technical system that is superior to the series system due to its more powerful (additional) sensors. If components are represented artificially, the closeness to reality suffers, but gains are made in terms of safety, reproducibility and observability.

Along with the option of creating the environment and the vehicle artificially, there are tools that use a virtual representation in the form of computer simulations. Here the two fields that combine the real and the virtual have a gray background, because strictly speaking they do not exist, because the task of sensors and actuators is to switch between virtual and real signals. A real radar sensor cannot sense a virtual environment, and a virtual converter cannot create real voltage.

However, what are possible are combinations of artificial and virtual environments and vehicles. Examples of this are provided by different concepts of vehicle-in-the-loop (ViL). To close the loop made up of actions and reactions of the environment and the vehicle, real components are mapped in the simulation in the form of models. Here either the sensors or actuators mentioned are stimulated, i.e. artificially instigated (examples of this are simulation-based videos as stimulation for camera systems or dynamometers as stimulation for drive actuators), or the testing tools directly simulate the power signals, such as the electromagnetic wave, and try to represent real effects of sensors and actuators in the simulation with the aid of models. For more information on this, see Bock [27] or Hendricks [28]. The use of models described calls the meaningfulness of these testing tools into question. To get valid results using such models, it must be verified that these models do not contain any impermissible simplifications; here impermissible is to be seen in the context of the function, and means that deviations from reality are only permissible below the tolerances of the function. However, if this validity has been verified, the testing tool enables greater safety during the test execution, as parts of the environment and the vehicle only encounter each other in the virtual world. Due to the virtual components, these testing tools are distinguished by greater variability, observability and reproducibility. From an economic perspective, this testing tool has the advantage of varying the virtual environment easily or representing the vehicle in a wide range of variants. An economic disadvantage could be the validation of the models (see below). An advantage

of this testing tool is the option, based on the simulated vehicle, of performing tests early on during the development.

The last level of abstraction represents the combination of a virtual vehicle and the virtual environment: The software-in-the-loop testing tool represents the closed control loop by modeling relevant components in the simulation. In contrast to the previous testing tools, the entire testing world is virtual. The tests are safe, variable, observable and reproducible; there is also the option of using this tool early on during the development. The economic advantage is provided by the hardware independence, as there is no connection to real time any more. The execution of the tests is only limited by the computer power; simulations can be run day and night, and also parallel on a large scale. On the other hand, there is the necessary closeness to reality of the virtual test world, and therefore of every individual model: Only when the validity of the models used can be verified are virtual tests sufficiently conclusive for a production release. Accordingly, for the economic consideration of simulation-based procedures, the validation of the models must be considered above all.

The same challenge exists for the use of formal methods. Mitsch [26] writes in this context: "We do (…) prove that collisions can never occur (as long as the robot system fits to the model)." This means that even for formal methods, the degree of reality of the models used determines the conclusiveness of the results. For example, a particular challenge that is therefore a focus of the research is the formalization of the uncertainties of sensors or the property of other road users.

The discussion relating to testing tools shows the potential to speed up the testing. With the aid of the artificially created environment and vehicle, test cases can be set up and executed specifically. Additionally, the virtual approach enables the tests to be speeded up and run in parallel, depending on the computer power used.

However, the discussion also shows that the validity of the tests, and therefore their conclusiveness, presents a challenge when artificial and virtual components are introduced.

21.7 Conclusion

Autonomous driving is distinguished in particular by the omission of the human supervisor of assisted or partially-automated systems, and the supervisor's ability to correct these systems. The metric consisting of real driving and driver's license test that enables a conclusion about the safety of automation levels currently present in series production, is no longer valid for autonomous driving. The resulting loss of the reduction in the test cases means that current test concepts are not suitable for economically assessing the safety of a new system such as autonomous driving. Adhering to current test concepts would involve an economically unjustifiable overhead, and would result in an "approval-trap" for autonomous driving. However, the authors see three approaches for avoiding this "approval-trap".

Firstly, the evolutionary approach, or alternatively the transformation (see Chap. 10), seems necessary, as only the step-by-step introduction along the different dimensions of speed, scenery and degree of automation enables existing components to be taken over and reduces the range of tasks for the following releases. Secondly, the necessary test cases must be reduced based on field experience and statistical procedures. The challenge here is the metric that allows a conclusion to be drawn about the safety of the system based on the completed test cases. Thirdly, alternative testing tools must be used alongside real driving. Here it is not expected to be able to do without real driving completely, because a verification of validity is required to move test cases to ViL, SiL and procedures that formally prove safety.

Finally, it must be stated that the challenges presented should not only be solved internally by the automobile industry. Even if test concepts are optimized for autonomous driving, there will not be 100 % safety. Vision Zero remains a vision for now, particularly in mixed operation with additional road users. With the first accident caused by an autonomous vehicle, at the latest, the previously issued release will be put to the test. Accordingly, the basis for the production release should be discussed publicly by all concerned and be designed transparently.

References

1. Felkai, R., Beiderwieden, A.: Schaffen allgemeiner Voraussetzungen der Projektabwicklung. In: Projektmanagement für technische Projekte, pp. 7-49. Springer Fachmedien Wiesbaden (2013)
2. DAS EUROPÄISCHE PARLAMENT UND DER RAT DER EUROPÄISCHEN UNION: RICHTLINIE 2007/46/EG (2007)
3. Reuter, A.: Produkthaftung Produkthaftung in Deutschland. In: Werdich, M. (ed.) FMEA - Einführung und Moderation, pp. 121-137. Vieweg + Teubner Verlag (2012)
4. Gasser, T.M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., Flemisch, F., Häcker, D., Hesse, T., Huber, W., Lotz, C., Maurer, M., Ruth-Schumacher, S., Schwarz, J., Vogt, W.: Rechtsfolgen zunehmender Fahrzeugautomatisierung. Gemeinsamer Schlussbericht der Projektgruppe. Berichte der Bundesanstalt für Strassenwesen - Fahrzeugtechnik (F), vol. 83. Wirtschaftsverl. NW Verl. für neue Wissenschaft, Bremerhaven (2012)
5. Ward, R.B.: Revisiting Heinrich's law. In: Chemeca. Quality of life through chemical engineering (2012)

6. Hydén, C.: The development of a method for traffic safety evaluation: The Swedish traffic conflicts technique, Lund Institute of Technology. Department of Traffic Planning and Engineering (1987)
7. Donner, E., Winkle, T., Walz, R., Schwarz, J.: RESPONSE 3 - Code of Practice für die Entwicklung, Validierung und Markteinführung von Fahrerassistenzsystemen. In: VDA Technischer Kongress 2007, pp. 231–241, Sindelfingen (2007)
8. Spanfelner, B., Richter, D., Ebel, S., Wilhelm, U., Branz, W., Patz, C.: Herausforderungen in der Anwendung der ISO26262 für Fahrerassistenzsysteme. Challenges in applying the ISO 26262 for driver assistance systems. In: Lehrstuhl für Fahrzeugtechnik, TU München (ed.) 5. Tagung Fahrerassistenz, München (2012)
9. Weitzel, A., Winner, H., Peng, C., Geyer, S., Lotz, F., Sefati, M.: Absicherungsstrategien für Fahrerassistenzsysteme mit Umfeldwahrnehmung, Forschungsbericht FE-Nr. 82.0546/2012 (zum Zeitpunkt der Manuskripterstellung nicht veröffentlicht)
10. Schuldt, F., Saust, F., Lichte, B., Maurer, M., Scholz, S.: Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen. In: AAET 2013
11. Horstmann, M.: Verflechtung von Test und Entwurf für eine verlässliche Entwicklung eingebetteter Systeme im Automobilbereich, TU Braunschweig (2005)
12. Daimler AG: Mercedes-Benz präsentiert in Genf Limousine und Coupé der neuen E-Klasse (2009)
13. Fach, M., Baumann, F., Breuer, J., May, A.: Bewertung der Beherrschbarkeit von Aktiven Sicherheits- und Fahrerassistenzsystemen an den Funktionsgrenzen. In: 26. VDI/VW-Gemeinschaftstagung Fahrerassistenz und Integrierte Sicherheit, 6./7. Oktober 2010 in Wolfsburg (2010)
14. Burgdorf, F.: Eine kunden- und lebenszyklusorientierte Produktfamilienabsicherung für die Automobilindustrie, KIT Scientific Publishing; Karlsruher Institut für Technologie (2010)
15. Wirtschaftskommission der Vereinten Nationen für Europa (UN/ECE): Regelung Nr. 13-H — Einheitliche Bedingungen für die Genehmigung von Personenkraftwagen hinsichtlich der Bremsen (2010)
16. Baake, U., Wüst, K., Maurer, M., Lutz, A.: Testing and simulation-based validation of ESP systems for vans. ATZ Worldw **116**(2), 30-35 (2014). doi: 10.1007/s38311-014-0021-6
17. Rasmussen, J.: Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. IEEE Transactions On Systems, Man, and Cybernetics **SMC-13**(3), 257–266 (1983)
18. Donges, E.: Fahrerverhaltensmodelle. In: Winner, Hakuli, Wolf (eds.) Handbuch Fahrerassistenzsysteme, pp. 15–23 (2011)
19. Ständer, T.: Eine modellbasierte Methode zur Objektivierung der Risikoanalyse nach ISO 26262, TU Braunschweig (2011)
20. Bahr, M., Sturzbecher, D.: Bewertungsgrundlagen zur Beurteilung der Fahrbefähigung bei der praktischen Fahrerlaubnisprüfung. In: Winner, H., Bruder, R. (eds.) 6. Darmstädter Kolloquium Mensch + Fahrzeug: Maßstäbe des sicheren Fahrens. Ergonomia (2013)
21. Statistisches Bundesamt (Destatis): Verkehrsunfälle - Fachserie 8 Reihe 7 (2012)
22. Winner, H.: Quo vadis, FAS? In: Winner, H., Hakuli, S., Lotz, F., Singer, C. (eds.) Handbuch Fahrerassistenzsysteme, 3rd edn. Vieweg-Teubner-Verlag (2015)
23. Glauner, P., Blumenstock, A., Haueis, M.: Effiziente Felderprobung von Fahrerassistenzsystemen. In: UNI DAS e.V (ed.) 8. Workshop Fahrerassistenzsysteme, pp. 5–14, Walting (2012)
24. Eckstein, L., Zlocki, A.: Safety Potential of ADAS – Combined Methods for an Effective Evaluation. ESV (2013)
25. Tatar, M., Mauss, J.: Systematic Test and Validation of Complex Embedded Systems. ERTS-2014, Toulouse, 5–7 (2014)

26. Mitsch, S., Ghorbal, K., Platzer, A.: On Provably Safe Obstacle Avoidance for Autonomous Robotic Ground Vehicles. Robotics Science and Systems (RSS), 2013. Accessed 27 June 2014
27. Bock, T.: Bewertung von Fahrerassistenzsystemen mittels der Vehicle in the Loop-Simulation. In: Winner, H., Hakuli, S., Wolf, G. (eds.) Handbuch Fahrerassistenzsysteme, pp. 76-83. Vieweg + Teubner Verlag (2012)
28. Hendriks, F., Tideman, M., Pelders, R., Bours, R., Liu, X. (eds.): Development tools for active safety systems: Prescan and VeHIL. Vehicular Electronics and Safety (ICVES), 2010 IEEE International Conference on (2010)