Article

Enhancing Acceptance and Trust in Automated Driving trough Virtual Experience on a Driving Simulator

Philipp Clement [1,2,*,†] , **Omar Veledar** [2,†] , **Clemens Könczöl** [3,†] , **Herbert Danzinger** [2] , **Markus Posch** [2] , **Arno Eichberger** [1] and **Georg Macher** [4]

1 Faculty of Mechanical Engineering and Economic Sciences, Institute of Automotive Engineering, Graz University of Technology, 8010 Graz, Austria; arno.eichberger@tugraz.at
2 AVL List GmbH, Hans-List-Platz 1, 8020 Graz, Austria; omar.veledar@avl.com (O.V.); herbert.danzinger@avl.com (H.D.); markus.posch@avl.com (M.P.)
3 Faculty of Natural Sciences, Institute of Psychology, University of Graz, 8010 Graz, Austria; clemens.koenczoel@uni-graz.at
4 Faculty of Electrical and Information Engineering, Institute of Technical Informatics, Graz University of Technology, 8010 Graz, Austria; georg.macher@tugraz.at
* Correspondence: clement@student.tugraz.at or philipp.clement@avl.com; Tel.: +43-699-1987-0820
† These authors contributed equally to this work.

Abstract: As vehicle driving evolves from human-controlled to autonomous, human–machine interaction ensures intuitive usage as well as the feedback from vehicle occupants to the machine for optimising controls. The feedback also improves understanding of the user satisfaction with the system behaviour, which is crucial for determining user trust and, hence, the acceptance of the new functionalities that aim to improve mobility solutions and increase road safety. Trust and acceptance are potentially the crucial parameters for determining the success of autonomous driving deployment in wider society. Hence, there is a need to define appropriate and measurable parameters to be able to quantify trust and acceptance in a physically safe environment using dependable methods. This study seeks to support technical developments and data gathering with psychology to determine the degree to which humans trust automated driving functionalities. The primary aim is to define if the usage of an advanced driving simulator can improve consumer trust and acceptance of driving automation through tailor-made studies. We also seek to measure significant differences in responses from different demographic groups. The study employs tailor-made driving scenarios to gather feedback on trust, usability and user workload of 55 participants monitoring the vehicle behaviour and environment during the automated drive. Participants' subjective ratings are gathered before and after the simulator session. Results show a significant increase in trust ensuing the exposure to the driving automation functionalities. We quantify this increase resulting from the usage of the driving simulator. Those less experienced with driving automation show a higher increase in trust and, therefore, profit more from the exercise. This appears to be linked to the demanded participant workload, as we establish a link between workload and trust. The findings provide a noteworthy contribution to quantifying the method of evaluating and ensuring user acceptance of driving automation. It is only through the increase of trust and consequent improvement of user acceptance that the introduction of the driving automation into wider society will be a guaranteed success.

Keywords: automated driving (AD); driving simulator; expression of trust; acceptance; simulator case study; NASA TLX; advanced driver assistant systems (ADAS); system usability scale; driving school

1. Introduction

Advanced driver assistant systems (ADAS) and autonomous driving (AD) are stepwise turning into reality. According to the Society of Automotive Engineers (SAE) this process is defined in five levels. In Figure 1 the levels are depict from conventional manual driving—no driving automation (SAE Level 0)—through conditional automated driving

(SAE Level 3) with the fallback-ready user to the full driving automation (SAE Level 5), which covers all situations everywhere [1].
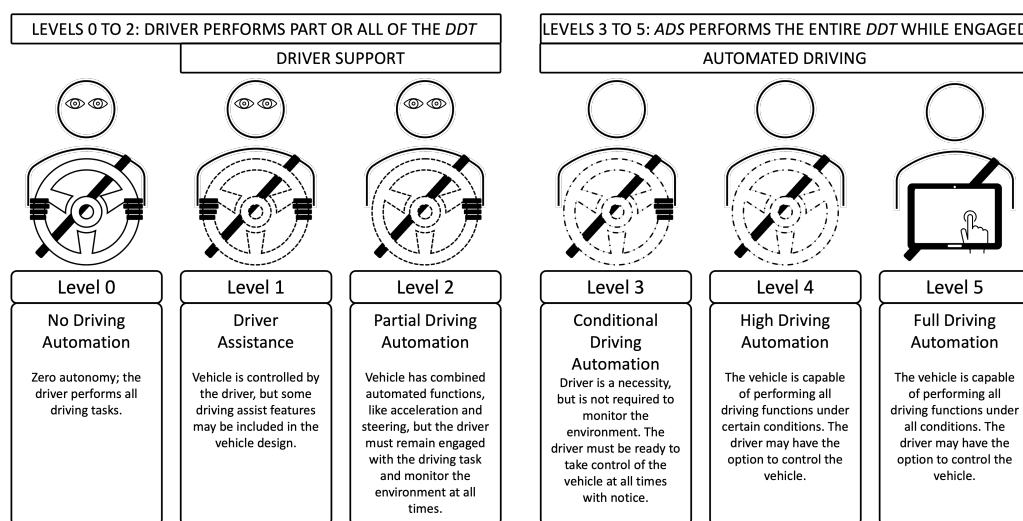

Figure 1. Society of Automotive Engineers (SAE) levels for driving automation following On-Road Automated Driving (ORAD) committee, SAE International [1] with Dynamic Driving Task (DDT).

In the intermediate steps, the human–machine interaction becomes an indispensable factor for the technological progress in the development [2,3], as it acts as a technological enabler that improves performance, safety, trustworthiness and comfort [4]. This is reflected by many studies concerning the interaction of humans with vehicle controls [5–7]. As there is a significant probability of harming the study participants, many studies are shifted from the real driving experience on the road into the virtual driving simulator environment, with particular attention on the selection of the appropriate simulator fidelity and validity [8]. This also turns out to be valid for future driver training programs with a potential expected to work even in countries with an already pretty low accident rate [9,10]. Moreover, simulators provide the opportunity to enhance training for safety-critical systems and situations [11–13]. The ongoing discussion of differences between real driving and the simulation has been overcome by the need for repeatability of the tests and the safety of the participants.

The future challenge on automated driving is that most of the time the task for the driver is to observe the environment and the vehicle performance without being actively involved. Then, occasionally in already near-critical situations the vehicle requests the driver to perform those critical situations immediately. These two contrary tasks challenge our society with the little knowledge about the technology and its interaction with human beings. Despite the potential benefits and influence of virtual training on the assessment of learning effects and trust in automation, there is a lack of sound research data on the effects of virtual training on those new systems. Besides study questionnaires, current research is ongoing to measure trust in automation in real-time scenarios, as in Azevedo Sá et al. [14].

1.1. The Technology Challenges

Multiple industrial sectors, including transportation and mobility, are experiencing a significant technological transformation. Considering the safety-critical nature of the technologies that drive this transformation, the automotive safety community has considerable interest in the development. However, the established safety engineering methods are somewhat failing to support these changes entirely. Thus, extensive effort is invested in developing safety engineering methods, safety design approaches and safety development processes that support this technological transformation [15].

AD is generally seen as one of the critical trends in the mobility sector and also one of the key components of the mentioned transformation [16]. To fulfil the anticipated AD

evolution, there is a demand for progress in supporting technologies (e.g., sensors and actuators), particularly in artificial intelligence (AI) (at the edge) based approaches [17] and further to a combined driver evaluation out of non-obstrusive monitoring sensors and the AI technology [2]. AI technologies are increasingly used for critical applications because these approaches exceed the current state of the art (e.g., recognising patterns and inferring relationships), creating a high demand for AI in terms of realising automated and autonomous driving functions [18]. The outcomes are expected to be technically viable and compelling for humans, hence enticing user acceptance of AD. A lot of effort is put into completing the challenges related to trustworthy AI solutions to cater for human stakeholders [19]. Trust and safety are of utmost importance as to gain acceptance of AD functions.

The key challenge in this context is that the established safety engineering processes and practices (e.g., described by ISO 26262 [20] and ISO/PAS 21448 [21] for the automotive domain) have been only successfully applied in conventional model-based system development. In many industry contexts, none of the available safety standards have defined processes that explicitly consider the specifics of AI-based approaches. These factors include the requirements on dataset collections, the definition of performance evaluation metrics, the AI architectures, and the handling of uncertainty [18].

The ultimate goal of safety-critical systems is to maximise the evidence of a positive risk balance. Conventional safety engineering processes apply model-based system development to establish structured and human-understandable arguments of the risks inherent to the system under development. The risk-driven safety engineering processes implement a system function with the required service quality (i.e., safety integrity level). For AI-based systems, it is challenging to provide such service quality metrics. For example, AI-based concepts build upon probabilistic modelling, including random variables and probability distributions, to model situations and events. As such, an AI function model returns a probability distribution as output for specific input.

In general, AI-based approaches depend on the data used to parameterise its function and the process of parameterising (called training or learning). The quality of the used dataset and the choice of the AI architecture directly influence the quality of the function. In contrast, human-programmed model-based functions typically return a specific result and not a probability distribution to a particular input. In the context of this work, this AI-related topic also reappears in relation to the establishment of trust and acceptance of automated driving in general, and the selection of virtual experiences in particular. It contributes to the establishment of trust and acceptance of such systems, with the expectation to open the gate for implementation of new AI functionality in future.

1.2. Trust and Acceptance

Trust in automation is a key determinant for the adoption of automated systems and their appropriate use [22] and along with other factors, has a significant impact on interest in using an automated vehicle [23]. Studies suggest a strong connection between trust in technical solutions and user acceptance on the same level [24–26]. The key challenges in terms of acceptance of fully or partially autonomous vehicles are balanced between the needed interaction with the technology, the associated benefits and the hidden risks. In particular, it is the level of required driver's awareness in combination with the human trust towards automated decisions in all conditions that are associated with acceptance [24]. For example, the need for continuous monitoring limits user acceptance as it diminishes the benefit of securing the driver with more freedom for side tasks. The additional open questions that could limit user acceptance are presented by the need for the drivers to take over vehicle control at emergency scenarios or extreme conditions or when the driving automation fails [27], unless there is a dependable fail-operational strategy in place to take over the vehicle control [28]. Most of the user acceptance studies analyse the effects of cognitive factors, but with a limited impact from social aspects [25], since the decision-making process is heavily influenced by opinions that are most important to the people

making these decisions [29] and their hedonistic motivations [30]. While that is empirically demonstrated [31], it is also expected that acceptance of the driving automation would be closely related to social influence but with a possibly different strength across different cultures, e.g., some cultures are likely to exert stronger on social influence than other ones [24]. Hence, there is a need for wider conclusive studies based on empirical data across demographic groups.

While acceptance is an important condition for the successful implementation of automated driving vehicles [32], recent research considers the relation between attitudes towards automated systems and their actual use [33]. Reduced acceptance level is also correlated to a lack of safety [34]. Furthermore, it is important to consider that currently AD has not yet been experienced by the majority of drivers [35]. Driving simulators are used to conquer the acceptance obstacle by determining the trust in AD functions and improvements of trust through experience while eliminating the need for complete vehicle prototypes and securing entire scenario repeatability [36]. It is the reproducible validation and demonstration of mature automated functions, their reliability, and safety that work towards securing societal acceptance [37]. Besides ensuring the durability of the technical solutions, increased user acceptance of the new automotive technologies crucially determines the sustainability of the consequent business implementation [38]. Sustainability is further supported through the trustworthiness of the solutions, which acts as an enabler for the development and implementation of an appropriate value creation strategy for maximisation of benefits amongst the engaging stakeholders [38].

In addition to focusing on the AD functions, the driving simulation potentially educates newly qualified drivers on unexpected critical situations on the road, which are generally unpredictable and not part of the driving license education. The potential of such training is evaluated by pre- and post-testing questionnaires for a broad spectrum of users, with attention to the low yearly driving experience group. There is a potential to increase acceptance and trust levels and situational experience through virtual training.

The consequent research questions are: can an advanced simulator (Section 2.4) experience improve trust in and the acceptance of automated driving, and are there any significant differences resulting from the driving simulations for different demographic groups (i.e., gender, age, experienced vs. inexperienced drivers, ADAS/AD experience)?

To answer these research questions we employ a unique combination of a state-of-the-art dynamic simulator in its characteristic measurement setup, the relevant scenarios targeting appropriate psychological assessment and a sizable study sample.

2. Materials and Methods

Simulators have the potential to raise the AD profile amongst the general population. To examine the impact of such exposure on a safe version of AD, a study comprising ten driving scenarios was designed to gather feedback on tailored questions from a defined sample group.

The expression of trust (EOT) questionnaire is the primary measurement tool in the present study to evaluate participant's subjective trust. The expression of trust is a modified version of the Trust into Automation questionnaire [39]. Each participant faces the questionnaire twice within the study to enable an initial estimate of trust and explore the change of trust in an AD system throughout the designed experiment. The questionnaire is used before the test as a baseline for each participant. Upon the test procedure, the participants face the same questionnaire again. This allows to measure the impact of such a simulator experience on the trust in the AD system.

The system usability scale (SUS) [40] is used to verify the suitability and usability of the simulator and the simulation within the testing environment, and the behaviour of the system itself. Furthermore, it indicates whether or not such a testing method is suitable for such a study.

A raw NASA TLX questionnaire [41,42] is used to research the workload of the participants. As the study does not request the participants to drive or react and control the

vehicle on their own, the NASA TLX reflects the workload of the participants monitoring the vehicle behaviour and environment during the automated drive. This effort is expected to be low due to the eliminated need for active driving tasks from the participants.

2.1. Sample

A total of 60 out of 112 registered participants participated in the test procedure. The selection focused on obtaining a well-distributed study sample. Out of those 60 participants, 55 completed the test successfully. The results from 7 participants were removed due to incomplete datasets for the questionnaires. This yielded 48 complete datasets for the analysis of the questionnaires of expression of trust, raw NASA TLX, and the system usability scale.

The sample is grouped by the participants' demographic factors in the defined categories, such as age, gender, yearly driving range, driver assistance experience, and educational level. The distribution is as follows:

- gender: 27 male, 21 female, 0 neutral;
- age group: 21 (18–29 years), 19 (30–45 years), 8 (46–65 years);
- mileage km/year: 8 (<5000 km/yr), 12 (5001–10,000 km/yr), 28 (>10,001 km/yr);
- ADAS experience: 22 (without pre-experience), 26 (with pre-experience);
- education: 13 below Bachelor level, 35 above Bachelor level—according to ISCED [43].

2.2. Study Materials

Three different questionnaires were used to evaluate trust, system usability, and workload to obtain subjective ratings of these relevant aspects. They gather the participants' feedback on their global perception of trust in automation and the use of AD systems. The questionnaires are the EOT, the raw NASA TLX, and the SUS. The questions are presented using a tablet computer allowing the subjects to provide their rating by screen tapping. The ratings may be corrected up to the point when the participant confirms the questionnaire as complete. The answer is recorded in a decimal format. To support the participants to put in their intended answer, the graphical appearance of the answer input area is supported by a color code, smileys, and written explanation. An example of the visualization is shown in Figure 2.
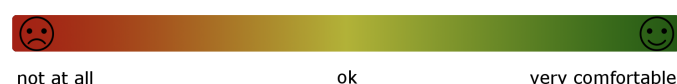
Q1 How comfortable did you feel?

Figure 2. Example of a question presented to the participants [44].

The EOT is a modified version of Helldin et al. [45] and relies on the original questionnaire on trust in automation from Jian et al. [39]. The questions assess user trust in AD functionality on a seven-point Likert scale, spreading from do not agree at all to agree completely. The questionnaire is presented before the participants get into the cockpit and start the test execution, and after they leave the cockpit. This double testing allows measuring changes in the participants' trust subject to their experience in the test environment. The following questions are used in this questionnaire:

Q1　I understand how the automated driving system works—its goals, actions, and results.
Q2　I would use the automated driving system if they were available in my own car.
Q3　I think the actions of the automated driving system will have a positive impact on my own driving.
Q4　I put my faith in the system.
Q5　I think the automated driving system will ensure safety while driving.
Q6　I think the automated driving system is reliable.
Q7　I can trust the automated driving system.

Each question reflects upon a specific part of the human trust in automation. In the present study, all single answers of a participant are summarised, and a mean value is built according to Equation (1) to reflect an overall change of the response, from before to after the experiment, where $Q0_p$ is the new built overall score per participant and A_{p_i} is the single response to one of the questions one to seven of one participant.

$$Q0_p = \frac{1}{n} \sum_{i=1}^{n} A_{p_i} \tag{1}$$

The mean value is treated as a new dependent variable and processed together with the demographic interactions in the analysis of variances and correlations.

After the test execution, the usability of the system is assessed by the SUS [46]. The ten-question questionnaire is adapted to fit the boundaries of the study. It also adopts the advice from Grier et al. [47]. The questions used are:

Q1 I think I would like to use this assistance system frequently.
Q2 I thought the assistance system was unnecessarily complex.
Q3 I thought the assistance system was easy to use.
Q4 I think I would need the assistance of a technical person to use this system.
Q5 My impression was that the different functions in this assistance system were well integrated.
Q6 I felt there was too much inconsistency in this assistance system.
Q7 I can imagine that most people would learn to use this assistance system very quickly.
Q8 I found that the assistance system was very cumbersome to use.
Q9 I found the use of the assistance system to be convincing.
Q10 I had to learn a lot of things before I could start using this assistance system.

The possible answers range from absolutely disagree to absolutely agree. The numerical values behind the answers reach from 0 to 4, with four representing the highest value the participant's answer can be. Therefore, for questions 1, 3, 5, 7, and 9, the values range from 0 to 4 and for the negatively formulated questions 2, 4, 6, 8, and 10, the values range from 4 to 0 and their polarity is reversed prior to the statistical analysis. For the ten questions, the maximum summed up result is 40 points. The summed value is further scaled by a factor of 2.5 to obtain a scale from 0 to 100 for each participant. A mean value for the system is then generated out of all participants' answers [48,49].

The third post-testing questionnaire is the NASA TLX [50], a standardised questionnaire on participants' perceived workload. For the actual purpose, the questions are presented as single questions without the weighting of the question pairs, also called "raw TLX". This was chosen as an appropriate cut-off in terms of the participant's timing. Therefore, an overall workload calculation is excluded in the data analysis procedure as the official rules can not be followed [50].

Q1 Mental demand. How mentally demanding was the task?
Q2 Physical demand. How physically demanding was the task?
Q3 Temporal demand. How hurried or rushed was the pace of the task?
Q4 Performance. How successful were you in accomplishing what was required of you?
Q5 Effort. How hard did you have to work to achieve your level of performance?
Q6 Frustration. How insecure, discouraged, irritable, stressed, and angry were you?

The answer options to those questions range from very low to very high with exception of question 4, which offers an answer scale from perfect to failed on a seven-point scale.

2.3. Experimental Procedure

Each participant followed a standardised experimental procedure also described in Clement et al. [44]. Figure 3 offers a top-level visualization of this procedure, starting with the introduction phase and continuing on to the post-testing phase. These two phases are the measurement phases for investigating the change of attitude due to the participant's exposure to AD in a moving driving simulator.

In phase one, the "introduction phase", the participants are introduced to the testing procedure, informed about the relevant data protection rules (according to the GDPR [51]) and the appropriate measurements. This is done by a psychologist to ensure consent and to provide a trained professional to deal with unforeseen circumstances. Upon the initial talk, the participants' expression of trust in AD and their prior experience is evaluated by a questionnaire. Participants have the ability to withdraw from the experiment in any phase.

In phase two, the "get prepared phase", the participants are equipped with the necessary sensors and connected with the database for recording. They are seated in the simulator for further instructions, which is the first time they see it. The participants do not see or meet each other to avoid any influence in this regard. They get informed about the detailed testing procedure and the equipment used for it, especially the simulator vehicle cockpit they are seated in and which they need to handle. Especially the interaction with the human–machine interface (HMI) is explained, as the information about disabled functionality is displayed in the drivers dashboard and interaction with the presented Scenario Specific Questionnaire on Trust (SSQT) questions is done on the central infotainment.

In phase three, the "get used to it phase", the participants are able to drive in the virtual environment on their own, five minutes without the moving hexa-pot platform and then five minutes with the activated movement platform. Test questionnaires of the SSQT are presented in the central infotainment dashboard to avoid uncertainties within the experiment.

In phase four, the "test execution phase", the actual exposure to the high automated driving scenarios takes place. The participants are driving trough the scenarios with the task to observe the behaviour of the vehicle and its reactions to the environment and further answer the SSQT to provide the required feedback. All ten scenarios and their consecutive SSQTs are processed automatically one after another.

In phase five, the "get cleared phase", the participants leave the cockpit, once all scenarios are completed successfully and all sensors are removed, so they are able to recover before they move on to phase six.

In phase six, the "post-testing phase", the same questionnaire as in phase one is presented to assess the participants' expression of trust in the AD after their simulator experience to evaluate the differences. Additionally, the NASA TLX [50] questionnaire is presented to the participants for workload measurements without a pairwise weighting. Furthermore, the SUS [40] is used to evaluate the subjective usability of the testing system.
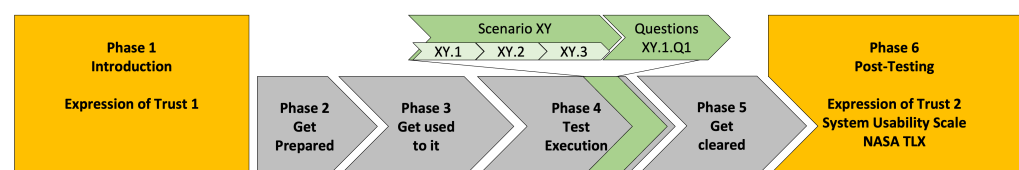

Figure 3. Experiment procedure for each individual participant. (XY represents the number of the scenario, XY.1 the first scene within the scenario XY, and XY.1.Q1 represents question one targeting the specific scene within the scenario XY).

2.4. Equipment and Techniques

For the virtual environment, an actual vehicle cockpit, resected of a real vehicle and modified with additional HMI displays and buttons, is mounted on a moving hexa-pod platform [52]. A 180° canvas displays the simulated scenery of a real road with three video projectors (4 K 100 Hz) for prevention against motion sickness. The platform is capable of performing movements in six degrees of freedom, three lateral directions with accelerating up to 6 m/s^2 in any direction and up to 1.5 m of movement, and rotating around the three axis as a result. The virtual environment is provided by Vires VTD [53], whereas the platform movement is calculated by a vehicle simulation model [54] for better accuracy. All components of the environment are synchronised by the co-simulation framework

Model.Connect [55]. The simulator is capable of performing all movements of normal driving conditions up to the aforementioned 6 m/s^2 accelerations [44].

The environment is regulated from a control (and preparation) room and is supported by a data server for data management. The automated data collection focuses on the participant's state, responses to the questionnaires, and the biometric data. The biometric sensor setup can be modified in accordance with the study design and the expected output. This study makes use of time-of-flight cameras, a chest belt for heart-rate measurements, a wristband for assessing skin conductivity and body temperature, and an eye-tracking device. The setup is characterised by the combination of the subjective psychological data assessment and the interactive subjective questionnaires.

2.5. Scenarios

Phase 4 of the testing procedure, shown in Figure 3, contains ten consecutive scenarios which are executed and standardised for each participant. The scenarios are designed to give the subject the impression of and experience with future AD. All scenarios contain potentially safety critical situations, which may influence participants' subjective perceived safety and therefore affect trust in automated driving systems. Even though participants do not get harmed in the simulated situations, they can neither predict the system's behaviour nor the final outcome. The simulated safety critical scenarios represent situations that may happen in real world driving with an estimated impact on trust and an influence on participants' subjective ratings. Each scenario is followed by an SSQT, supported by a reminding video of the action to be evaluated. The scenarios are scheduled automatically from the test automation service once the SSQT is completed. The ten scenarios are divided into the following 4 clusters (scenarios with similar use case [56]), which are also shown in Figure 4.

Simulated safety critical situations:

C1 Take-over manoeuvres with function decrease (2 scenarios);
C2 Emergency brake with bad weather conditions (2 scenarios);
C3 Narrow road with a child crossing (3 scenarios);
C4 Highway with construction site ahead (3 scenarios).

Each scenario is conducted in high driving automation mode (SAE level 4—no need for driver intervention, but still possible) [1], see also Figure 1, while the participants were tasked with observing the vehicle behaviour and the environment and providing their subjective feedback. The scenario clusters are designed in the same environment [56]. Their AD parameters differ so that two driving modes are available: sporty or comfortable. Compared to the comfortable configuration, the sporty configuration allows a shorter time to collision and higher de-/acceleration rates as well as smaller distances to follow or to stop. Beyond the description of the scenarios, they are published for all details in [57].

Cluster one describes a constant drive with 100 km/h approaching a vehicle with 70 km/h on a straight one-way two-lane road in an urban area. The driver is informed about the high driving automation mode in the HMI. As all parameters like speed difference, other traffic and range of vision for an automated take-over are met, the vehicle controller performs an automated take-over of the other vehicle. After the take-over, the ego vehicle automatically changes back to the initial lane and continues to drive with 100 km/h. Next, it is passed by another vehicle which splashes dirt on the ego vehicle's sensor setup. The driver gets informed about a decrease in take-over functionality. The next vehicle approached by the ego vehicle is driving with 70 km/h. As the sensor setup is not recovered yet, no automated take-over can be performed and the vehicle follows the slower vehicle in the lane in front of the ego vehicle. After a predefined time (10 s) of sensor cleaning, the take-over functionality is available again and displayed to the driver. As all mandatory conditions are met, another automated take-over is performed in the same manner as the first one. After changing back to the initial lane, the drive continues with 100 km/h and the scenario ends. This scenario is performed twice, once with each controller configuration.
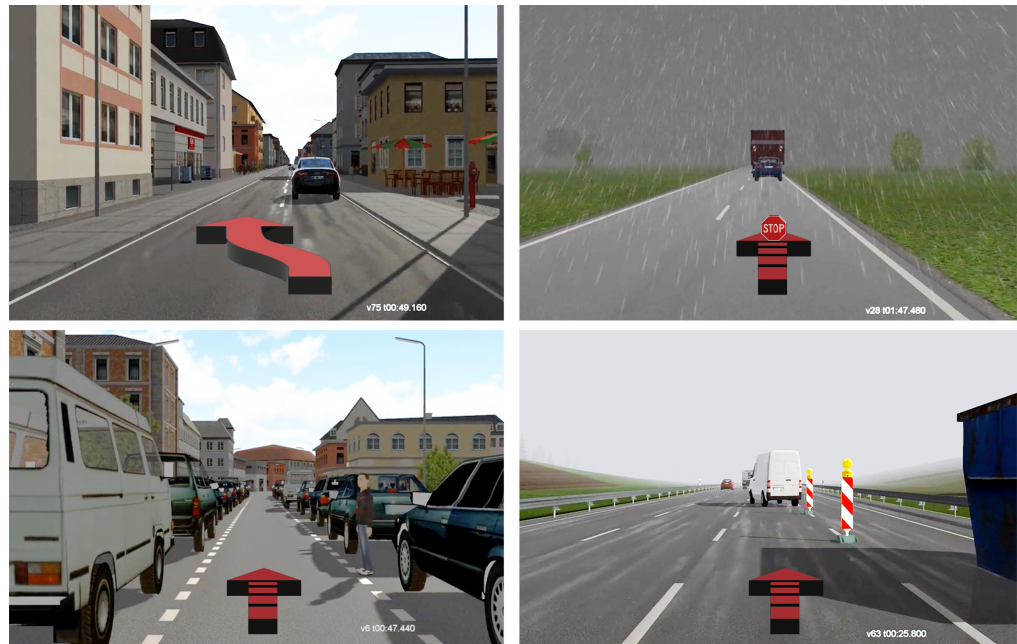
Figure 4. Pictures of critical events of the four scenario clusters: (**top left**) C1—at the beginning of the second take-over; (**top right**) C2—appearance of the stopped truck in the fog; (**bottom left**) C3—appearance of the child between the parked vehicles; (**bottom right**) C4—cut-in of the delivery van from the construction site with nearly zero speed.

Cluster two is described by the ego vehicle starting off and following another vehicle driving with 50 km/h in high driving automation mode on a straight two-way road. The speed remains constant and appropriate to the road conditions. After a defined time of 40 seconds following the vehicle ahead, the weather conditions deteriorate, as the appearing fog limits visibility. Due to the reduced visibility of the sensors, the vehicle's speed is automatically reduced by the automated driving controller to 30 km/h and the driver is informed about the issue via the HMI. The vehicle in front drives faster and disappears in the fog. As the drive continues, the vehicle ahead suddenly appears in the fog and seems to be stopping behind a truck that has already stopped in the fog. The automated driving system handles the emergency brake situation and automatically stops behind the vehicle in a suitable distance. Due to a higher speed, the *sporty* controller setting creates a more critical situation and a shorter stopping distance. No accident occurs. This scenario is performed twice, once with each controller configuration.

The three scenarios of cluster three are characterised by a high driving automation low speed drive in an urban residential area with parked vehicles. The road is straight. The ego vehicle approaches a narrow lane with vehicles parked on both sides of the one-way road with 30 km/h. The ego vehicle automatically changes lanes to the free middle lane and the speed is reduced further to 15 km/h due to reduced sensor range between the parked vehicles. After some driving time between the parked vehicles, suddenly a child appears between two parked vehicles on the right and crosses the road just ahead of the ego vehicle. The automated driving controller reacts to the situation and nearly stops <5 km/h the ego vehicle in front of the child. After the situation is cleared and the child leaves the scene, the controller automatically continues driving. This scenario is performed three times and the stopping distance and the deceleration values differ between the scenarios depending on controller configurations (1 & 3 comfortable and 2 sporty).

Cluster four implements three evolving scenarios. The consecutive scenario always extends the previous one by adding lousy weather in the second repetition and a delivery van cutting into the driving line in the third repetition, while the controller for the high driving automation is always set on comfortable mode. All scenarios are driven in high driving automation mode. The basic/first scenario is defined by driving on a three-lane

motorway on the first lane with 100 km/h, which is then closed due to a construction site ahead with a speed limitation of 80 km/h. The automated driving controller automatically changes from the first to the second lane as all conditions for an automated lane change are met. The speed is reduced as required by the traffic signs and the construction site is passed successfully. After the construction site the speed is increased back to 100 km/h and the scenario ends. In the second scenario, the first scenario is repeated under poor weather conditions and reduces the visibility. The rest of the scenario remains unchanged, while in the third scenario, the weather conditions from the second scenario remain the same and an additional critical situation is provoked by a delivery van leaving the construction site right in front of the ego vehicle. The automated driving controller performs an emergency brake and reduces the speed from 80 km/h to 18 km/h to avoid a collision. The distance to the delivery van is reduced down to 2 m and creates a critical situation at higher speeds. The delivery van accelerates slowly as is typical for the vehicle type. After the construction site, the delivery van changes to the first lane. The ego vehicle lane is cleared and the automated driving controller accelerates to the initial speed of 100 km/h and the scenario ends.

3. Results

The results focus on the participants' subjective ratings on the questionnaires, measuring system usability, trust and workload. The results are divided into subchapters for each questionnaire. The subjective ratings are analysed using multifactorial analysis of variance (MANOVA) with repeated measurements with a significance level of $p_{unc} \leq 0.05$ of the null-hypothesis. The p-value index indicates the specific analysis. So p_{int} stands for the interaction of the relevant groups and the factor time, p_t for the factor time, and p_{gr} for the group factors. Moreover, Pearson correlations were calculated to examine the correlations between the different factors. The method used to process the gathered results is already made available in Clement et al. [44].

3.1. Expression of Trust

The evaluation of the expression of trust uses the mean value (Q0) to analyse the participants' trust within the sample throughout the experiment.

A tendency towards a significant interaction can be seen in combination with ADAS pre-experience ($F = 2.308$, $p_{int} = 0.136$, $\eta^2 = 0.048$). The group with ADAS pre-experience shows a lower increase in the overall response than the group without ADAS pre-experience. This can be seen in Figure 5 on the very left side at the Q0 values. The ADAS pre-experience group's results show that the ratings before and after the experiment are at a similar level but slightly increase for the post-test results. For the factor time, the rating changes highly significant from 4.58 to 4.98 point ($F = 10.160$, $p_t = 0.003$, $\eta^2 = 0.181$). For the factor ADAS pre-experience, a tendency towards a significant difference can be seen ($F = 2.940$, $p_t = 0.093$, $\eta^2 = 0.060$).

The mean values resulting from the Expression of Trust (EOT) questionnaire before and after the experiment are shown in Figure 5, including their standard deviation. The results are split into two key groups, with and without ADAS pre-experience. The figure shows this for the mean value (Q0) and for the questions Q1 to Q7. The details are in Table 1. The single questions show various noticeable and significant differences. In Q1 "I understand how the system works—its goals, actions and output", the interaction ($F = 10.619$, $p_{int} = 0.002$, $\eta^2 = 0.188$) and the factor ADAS pre-experience itself ($F = 6.462$, $p_{gr} = 0.014$, $\eta^2 = 0.123$) show a significant difference. Further, in Q3, the interaction reveals a significant difference between the two groups of the factor ($F = 6.384$, $p_{int} = 0.015$, $\eta^2 = 0.122$), as indicated in the empirical analysis above. In Q5, the factor pre-experience itself reveals two nearly significant different levels of the answers ($F = 3.962$, $p_{gr} = 0.052$, $\eta^2 = 0.079$), also indicated by the different empirical levels of the group answers. Ratings of participants without ADAS pre-experience are higher for all answers in the post-test questionnaire compared to the

pre-test questionnaire. The ratings provided by the participants with ADAS pre-experience are less consistent with a higher variance before and after the experiment.
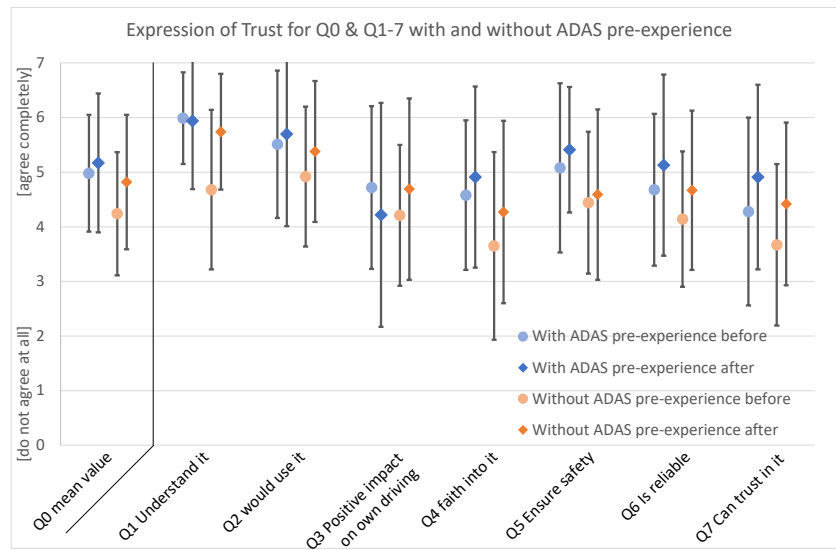

Figure 5. Expression of trust for question 0 (Q0—mean value) and detailed question 1–7 (Q1–7) with and without ADAS pre-experience, before and after the test.

Table 1. Mean values (m), standard deviation (sd) and ANOVA-results for ADAS pre-experience x Time.

| | with ADAS Pre-Experience | | | | without ADAS Pre-Experience | | | | ANOVA | |
	(m) before (sd)		(m) after (sd)		(m) before (sd)		(m) after (sd)		ADAS ($F/p_{gr}/\eta^2$)	Interaction ($F/p_{int}/\eta^2$)
Q0	4.98	(1.07)	5.17	(1.27)	4.24	(1.13)	4.82	(1.23)	*(2.940 / 0.093 / 0.060)*	*(2.308 / 0.135 / 0.048)*
Q1	5.99	(0.84)	5.94	(1.25)	4.68	(1.46)	5.74	(1.06)	*(6.462 / 0.014 / 0.123)*	*(10.619 / 0.002 / 0.188)*
Q2	5.51	(1.35)	5.70	(1.69)	4.92	(1.28)	5.38	(1.29)	*(4.535 / 0.217 / 0.033)*	*(0.593 / 0.445 / 0.013)*
Q3	4.72	(1.49)	4.22	(2.05)	4.21	(1.29)	4.69	(1.66)	*(0.001 / 0.973 / 0.000)*	*(6.384 / 0.015 / 0.122)*
Q4	4.58	(1.37)	4.91	(1.66)	3.65	(1.72)	4.27	(1.67)	*(3.382 / 0.072 / 0.068)*	*(0.548 / 0.463 / 0.012)*
Q5	5.08	(1.55)	5.41	(1.15)	4.44	(1.30)	4.59	(1.56)	*(3.962 / 0.052 / 0.079)*	*(0.260 / 0.612 / 0.006)*
Q6	4.68	(1.39)	5.13	(1.66)	4.14	(1.24)	4.67	(1.46)	*(1.749 / 0.192 / 0.037)*	*(0.068 / 0.795 / 0.001)*
Q7	4.28	(1.72)	4.91	(1.69)	3.67	(1.48)	4.42	(1.49)	*(1.885 / 0.176 / 0.039)*	*(0.077 / 0.783 / 0.002)*

Note: (F ... F-value, p_{gr} ... p-value for Group ADAS pre-experience, p_{int} ... p-value for interaction, η^2 ... effect size) of the expression of trust for Q0 to Q7 for the factors ADAS pre-experience (with/without) and time of measurement (before/after).

The factor gender is illustrated in Figure 6 regarding the mean value (Q0) and the single questions Q1 to Q7. Results show a tendency towards the same significant difference for time of measurement (before/after) for both groups, male and female. However, although they reflect the overall change of trust through the simulator session, they do not show a significant interaction regarding differences in Q0 ($F = 1.733$, $p_{int} = 0.195$, $\eta^2 = 0.036$). The global picture shows that the female group tended to rate the system lower than the male group before the experiment. After the experiment, the increase of the female group's rating was higher than the one of the male group but still did not surpass the level of the male group.

In Table 2, results for the analysis of variance for the expression of trust are given for the interaction between the factor time and the factor gender as well as the main effect for the factor gender. The main effect regarding the factor time is the same as mentioned in the analysis of ADAS pre-experience and not outlined separately. Significant differences can be seen for the interaction in Q4 ($F = 5.230$, $p_{int} = 0.026$, $\eta^2 = 0.102$) and for the factor gender in Q1 ($F = 5.510$, $p_{gr} = 0.023$, $\eta^2 = 0.107$).
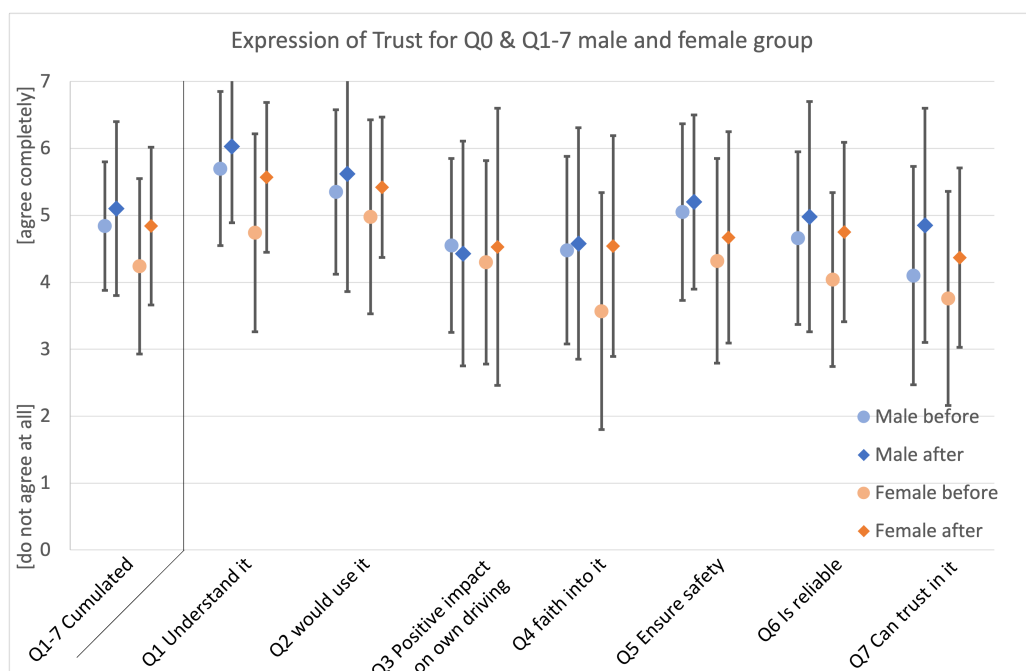
Figure 6. Expression of trust for question 0 (Q0—mean value) and detailed question 1–7 (Q1–7) for male and female groups, before and after the test.

Table 2. Mean values (m), standard deviation (sd) and ANOVA-results for Gender x Time.

| | Male | | | | Female | | | | ANOVA | |
	(m) before (sd)		(m) after (sd)		(m) before (sd)		(m) after (sd)		Gender $(F/p_{gr}/\eta^2)$	Interaction $(F/p_{int}/\eta^2)$
Q0	4.84	(0.96)	5.10	(1.30)	4.24	(1.31)	4.84	(1.18)	*(1.801 / 0.186 / 0.038)*	*(1.733 / 0.194 / 0.036)*
Q1	5.70	(1.15)	6.03	(1.14)	4.74	(1.48)	5.57	(1.12)	*(5.510 / 0.023 / 0.107)*	*(1.836 / 0.182 / 0.038)*
Q2	5.35	(1.23)	5.62	(1.76)	4.98	(1.45)	5.42	(1.05)	*(0.597 / 0.443 / 0.013)*	*(0.223 / 0.639 / 0.005)*
Q3	4.55	(1.30)	4.43	(1.68)	4.30	(1.52)	4.53	(2.07)	*(0.032 / 0.859 / 0.001)*	*(0.701 / 0.406 / 0.015)*
Q4	4.48	(1.40)	4.58	(1.73)	3.57	(1.77)	4.54	(1.65)	*(1.193 / 0.280 / 0.025)*	*(5.230 / 0.026 / 0.102)*
Q5	5.05	(1.32)	5.20	(1.30)	4.32	(1.53)	4.67	(1.58)	*(2.867 / 0.097 / 0.059)*	*(0.344 / 0.560 / 0.007)*
Q6	4.66	(1.29)	4.98	(1.72)	4.04	(1.30)	4.75	(1.34)	*(1.256 / 0.268 / 0.027)*	*(1.377 / 0.246 / 0.029)*
Q7	4.10	(1.63)	4.85	(1.75)	3.76	(1.60)	4.37	(1.34)	*(1.001 / 0.322 / 0.021)*	*(0.089 / 0.767 / 0.002)*

Note: (F ... F-value, p_{gr} ... p-value for group gender, p_{int} ... p-value for interaction, η^2 ... effect size) of the expression of trust for Q0 to Q7 for the factors gender (m/f) and time of measurement (before/after).

Analysis for the remaining demographic subgroups on the mean value (Q0) show no significant differences for the factor age ($F = 0.020$, $p_{gr} = 0.980$, $\eta^2 = 0.001$) and for the interaction with the factor time ($F = 0.183$, $p_{int} = 0.834$, $\eta^2 = 0.008$). For the factor driving experience, the values show no significant differences ($F = 1.075$, $p_{gr} = 0.350$, $\eta^2 = 0.046$), same as for the interaction with the factor time ($F = 1.045$, $p_{int} = 0.360$, $\eta^2 = 0.044$).

3.2. NASA TLX

The NASA TLX was performed as a raw test without the weighting of the questions in pairs. The evaluation of the questionnaire reveals that the factor age group ($F = 3.481$, $p_{gr} = 0.039$) had a significant influence on participants' ratings of Q5. The group of elderly participants and those with fewer kilometers per year report more effort to accomplish their performance. Moreover, the factor driving experience ($F = 4.278$, $p_{gr} = 0.019$) had a significant influence on participants' ratings of Q6. These factors as well as the factor ADAS are depicted in Figure 7. There were no significant interactions reported in the two-way analysis of variance as shown in Table 3. The differences are noteworthy for Q5. For the group with less than 5000 km/year, a descriptive analysis shows lower success for the questions Q4, Q5, and Q6.

Table 3. Results of the two-way ANOVA analysis for the NASA TLX.

	Q1	Q2	Q3	Q4	Q5	Q6
Unifactorial Analysis					Age Group *	Driving Exp *
(A) Age group vs. (B) Driving Exp.	(A vs. B)	(A vs. B)	(A * vs. B)	(A ** vs. B)	(A ** vs. B)	(A ** vs. B)
(A) ADAS Exp. vs. (B) Driving Exp.	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B *)
(A) Age Group vs. (B) ADAS Exp.	(A* vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)
(A) Gender vs. (B) ADAS Exp.	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)
(A) Gender vs. (B) Age Group	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B *)	(A vs. B)
(A) Gender vs. (B) Driving Exp.	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B)	(A vs. B *)

Note: ** indicates $p \leq 0.01$, and * indicates $p \leq 0.05$ X. * indicates a significant difference for the factor, and (X vs. X)* indicates a significant interaction.
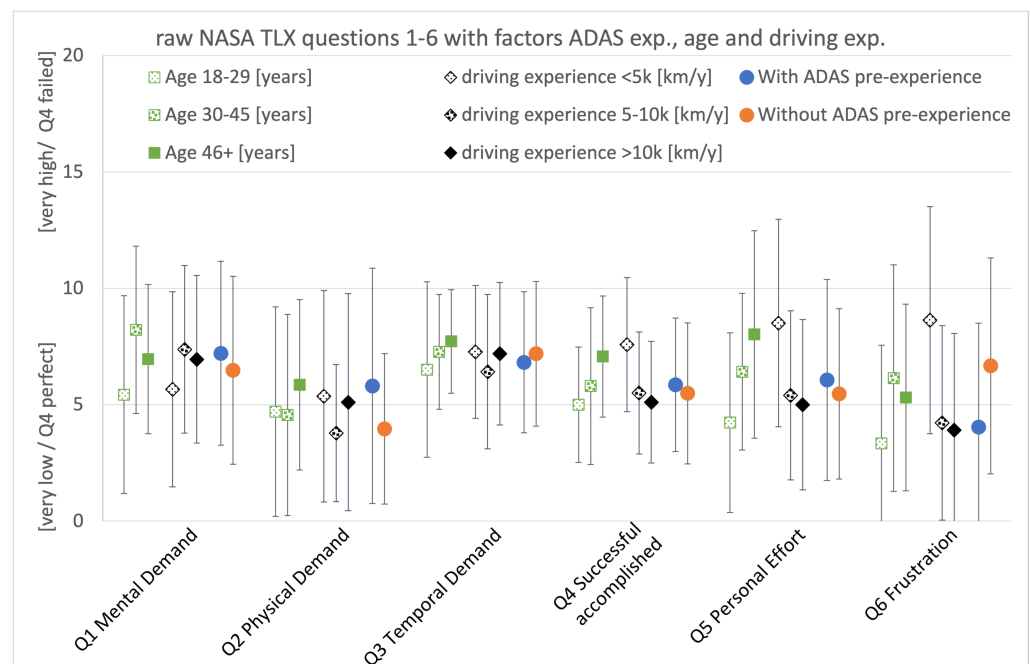

Figure 7. NASA TLX for questions 1–6 (Q1–6) for the groups with and without ADAS pre-experience, the groups of driving experience per year, and the age groups.

3.3. System Usability Scale

The system usability evaluated for each demographic sample group shows results around the overall value of 77.49 out of 100. The value is generated from the sum of all single questions for each subject multiplied by 2.5 to correspond to the target maximum scale of 100 according to Rauer [49]. Figure 8 shows the results for each demographic subgroup within the participants. The groups are all in a similar range and distribution.

A two-way analysis of variance revealed only one significant interaction of the age group with the driving experience (F = 4.347, p_{gr} = 0.005). This may be explained by a small number of participants in the age group of 46+ years, which results in a biased *p*-value sensitive to a distribution based on small sample size, even though according to Bangor et al. [46], the system used reflects an overall score in the 3rd quartile, which is acceptable and in terms of rating between good and excellent for the whole sample.
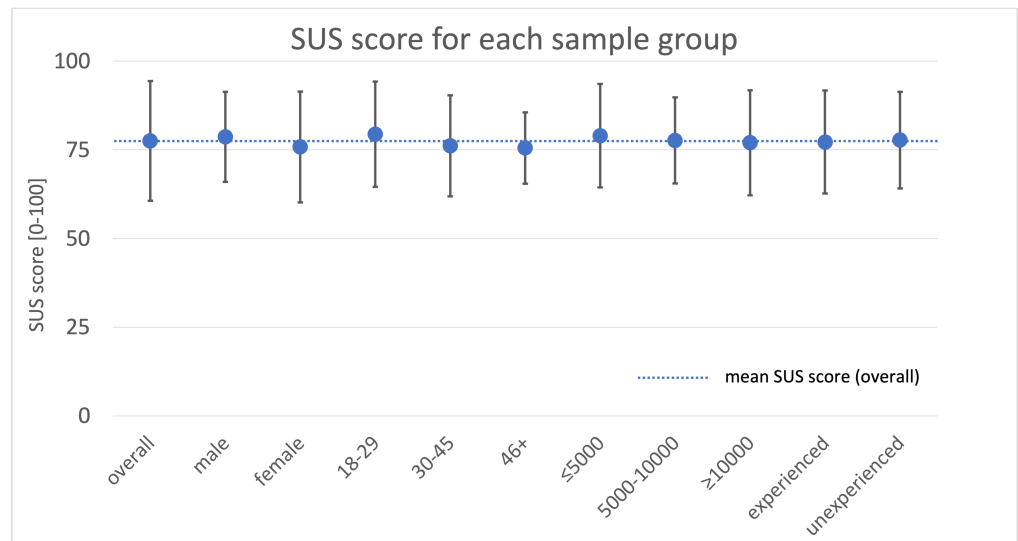
Figure 8. SUS score for the demographic sample groups.

The results for single questions are shown in Figure 9. Since the questions are positively and negatively polarised, the negative ones were inverted prior to the analysis. Results suggest an easy-to-use and quickly adjustable system with no unnecessary complexities. Its usage is not cumbersome and it demands no assistance to be used. The results also reveal that the system has some inconsistency (2.67 out of 4) and is not fully convincing (2.64 out of 4) to the overall sample, but still far better than the 50% quantile.
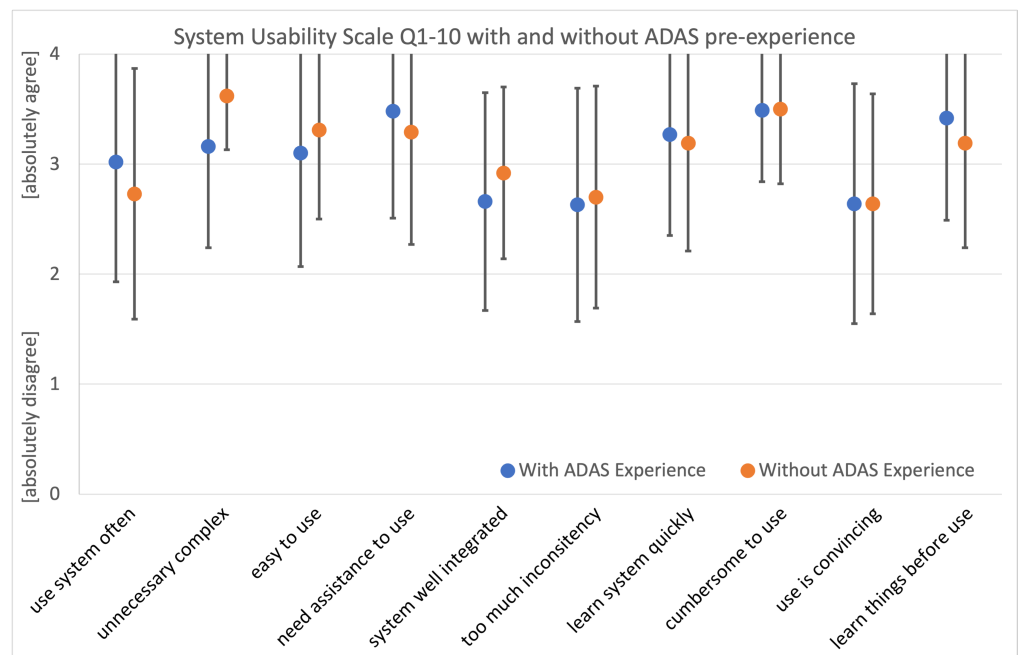
Figure 9. SUS scores for the single questions for the groups with and without ADAS pre-experience, inverted for negative questions.

*3.4. Correlations between Workload, Trust and Usability*

To identify the correlation between the participants' ratings, Pearson correlations between all questionnaires and single items were calculated. In Table 4, an overview of the major calculations is given using the mean values (Q0) for the SUS and the EOT, as they have higher reliability due to being built on the sub-questions of the related questionnaires and therefore get more information into the variance. The values show the correlation r and their significance.

Table 4. Pearson correlations between system usability scale (SUS), expression of trust (EOT) and NASA TLX items.

		1	2	3	4	5	6	7	8	9
1	**SUS_Q0**	-	0.37 *	0.63 ***	−0.47 **	−0.05	−0.40 **	−0.40 **	−0.17	−0.46 ***
2	**EOT_BEFORE_Q0**		-	0.75 ***	−0.14	0.21	−0.07	−0.23	−0.21	−0.31*
3	**EOT_AFTER_Q0**			-	−0.29	0.03	−0.16	−0.31 *	−0.06	−0.37 *
4	**NASA_Q1 (mental demand)**				-	0.48 **	0.24	0.22	0.41 **	0.28
5	**NASA_Q2 (physical demand)**					-	0.06	0.03	0.16	0.07
6	**NASA_Q3 (temporal demand)**						-	0.14	0.22	0.40 **
7	**NASA_Q4 (performance)**							-	0.38 **	0.42 **
8	**NASA_Q5 (effort)**								-	0.35 **
9	**NASA_Q6 (frustration)**									-

Note: *** indicates $p \leq 0.001$, ** indicates $p \leq 0.01$, and * indicates $p \leq 0.05$.

The results reveal the following significant findings. SUS (Q0) and EOT before (Q0) were found to be highly significant and moderately positively correlated ($r = 0.37$, $p = 0.01$). SUS (Q0) and EOT after (Q0) were found to be highly significant and highly positively correlated ($r = 0.63$, $p = 0.00$). EOT before (Q0) and EOT after (Q0) were found to be highly significant and highly positively correlated ($r = 0.75$, $p = 0.00$). Scatter plots of these correlations can be found in Figure 10. These positive correlations imply that participants who rated the system usability low also rated the expression of trust (before and after) lower, and those who rated the system usability higher also have a higher expression of trust. The correlation between SUS and EOT after is higher than SUS and EOT before. Moreover, significant correlations between several NASA TLX single items and the SUS (Q0), the EOT (Q0) before, and the EOT (Q0) after were found as illustrated in Table 4.
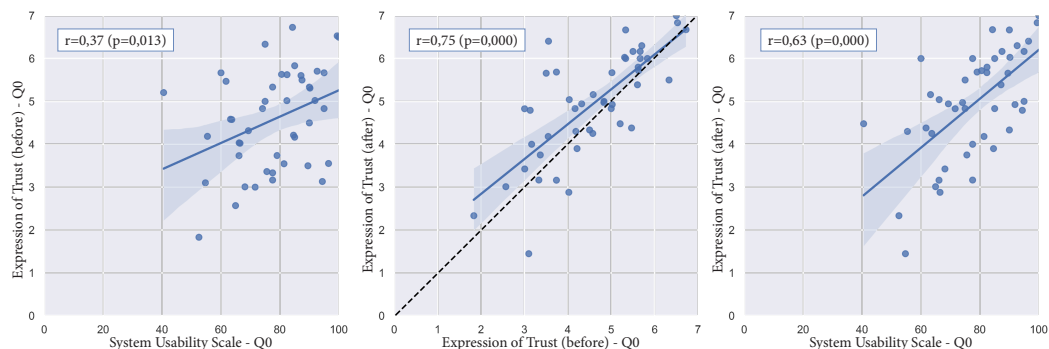
Figure 10. Pearson correlations for the combination of the accumulated questions Q0 with best fitted line. (r ... Pearson correlation coefficient, p ... significance of the correlation).

Results that reveal significant correlations between NASA TLX and SUS as well as NASA TLX and EOT after, they are consistently negatively correlated. This means the higher participants rate the workload, the lower they rate the usability and trust. These results reflect a high internal validity of the experiment. Furthermore, these correlations suggest a high validity of the applied questionnaires, as results point to a consistent and contextual evaluation within the measurement tools.

To check for spurious correlations and improve understanding of the correlations, especially regarding correlations' linearity and data outliers, the correlations are visually analysed using a scatter plot and the correlation line. The major correlations are depicted in Figure 10. The scatter plots do not show a tendency for a non-linearity and are well scattered around the correlation line. A visual comparison of the expression of trust (Q0) before and after the experiment reveals a general increase of trust as there are more data points above the 45° dotted line. A point on the dotted line would mean an equal rating in Q0 before and after the experiment and indicates no change throughout the experiment. A data-point

below the dotted line indicates a decrease of trust, and a data-point above it means an increase of trust. This approach supports the findings of the analysis of variance reported in Section 3.1. Apart from the inter-questionnaire correlations, the highest correlation and significance levels relate to the system usability scale with questions SUS Q1 (would use the system frequently) and SUS Q9 (found the system to be convincing) in correlation to the post-testing measurement of the expression of trust with questions Q2, Q4, Q6, and Q7.

4. Discussion

We quantified the increase of trust in the "AD system under test" by exposing participants to a driving simulator and evaluating their subjective perception before and after that experiment session. As expressed by the mean value (Q0) of the Expression of Trust questionnaire, the increase of trust was significant for all groups between the two points of measurement. This study shows an appropriate way to increase and evaluate trust in a simulator study.

Participants with ADAS pre-experience entered the study with higher confidence in such systems compared to the group without ADAS pre-experience, as depicted in Figure 5. Despite the high starting confidence level, the simulator session increases their trust in the AD system, indicating a high level of validity between the driving simulation and the real-world experience. There is a significant increase in trust for both groups. The group without ADAS pre-experience shows a much higher growth than the already pre-experienced. Besides the single questions Q1 and Q3 of the EOT, no significant difference can be seen between the groups. Q1 hints that the pre-experienced group could not gain further understanding, whereas the inexperienced group could significantly increase their knowledge of such a system. We found that pre-experienced participants might see a negative impact on their driving style the more they learn about the possibilities of automated driving systems and their driving behaviour in safety-critical situations, as a decrease in the Q3 was found. In contrast, the group without ADAS pre-experience sees a positive impact on their driving style, as they express a willingness to use the system in the future.

The analysis of demographic differences and similarities between genders shows the same tendency towards gaining trust (Q0) in the system through the simulator experience. The subgroups express a similar behaviour, besides Q1, with only one significant difference in the interaction of the groups and the time of measurement, which reflects the faith in the system shown in Figure 6. Female participants are more sceptical about their faith in the system before the simulator session compared to the male group. The simulator session affected their attitude towards the applied automated driving systems, i.e., their trust increased, reaching a similar level as the male participants. This effect can also be observed as a tendency in all other questions, even though there is no significant difference. This trend can also be seen within a descriptive analysis. It may reflect that the female group was more sceptical before the experiment but reached a level similar to one of the male participants after the experiment.

The NASA TLX analysis reveals an overall expected low workload of the participants. As the questionnaire is provided after the complete test sequence, it reflects the workload of the entire simulator session and does not differentiate between the single scenarios. There are significant negative correlations between the NASA TLX and both the system usability scale and the expression of trust. Since both questionnaires measure similar aspects with different purposes, the effect can be assumed as valid. The negative correlations between the trust in the system and the workload throughout the simulator session are significant, meaning that a higher workload correlates with less trust within the participants. The participants' ratings were below the 50% line for the overall workload evaluation, which suggests a low workload within the simulator session. This was expected, as full AD minimised the physical and mental demands. Nevertheless, some subgroups showed more strain within the experiment. As shown in Figure 7, the groups of higher ages and lower yearly mileage had a lower success rate, a higher personal effort and a higher frustration

after the simulator session. Hence, there are only tendentially significant differences between the groups.

The participants' ratings within the System Usability Scale (SUS) show a usability that can be seen as good to excellent. As already discussed in Bangor et al. [46] and Brooke [48], this method is suitable for evaluating the overall system usability. We hypothesise that the participant knowledge of the early developmental state of the system may have influenced their evaluations. With a mean of around 77.5 points out of 100, the response indicates a very good system although it also implies that following a normal distribution, around 50% of the participants rated the system lower than that and therefore only good instead of very good. The assessment of the mean and standard deviation of all demographic subgroups reveals that they are all on a similar level regarding their usability evaluation. This points to a neutral, not too complicated and realistic experimental design, which may be seen as an essential prerequisite for reliable results.

Regarding the Pearson correlation with the combination of both the pre- and post-questionnaires for the expression of trust, one can derive that smaller trust in the initial level results in a higher increase of trust after the simulator session. This is expressed in the middle of Figure 10 by the fitted line being above the 45° line on the left side while touching the 45° line at the top end, which may also reflect a ceiling effect as the scale is limited to 0–7 and may not be extended in the post-session questionnaire. An increase from a higher level may not be measured that accurately. A comparison of the pre- and post-correlations with the system usability shows a more homogeneous rating in the post-evaluation. The correlation with the pre-questionnaire is smaller ($r = 0.37$, $p = 0.013$) reflecting that the system's behaviour is unknown prior to the simulator session. Nevertheless, the lower part of the system usability scale shows a slight decrease in the expression of trust. In contrast, the upper part did increase noticeable in a clockwise turn of the fitted line in the left and right part of Figure 10.

Both correlations suggest different clusters, one below and one above the fitted line of Figure 10. An analysis of the correlations with a separation into the different demographic groups reveals noticeable findings which can be seen in Figure 11. It shows that the group without ADAS pre-experience seems to be located closer to the cluster below the fitted line than the group with ADAS pre-experience. It also reveals that the trust of the group with a low annual driving experience increases its mean ratings in trust throughout the simulator session and has the highest intern group delta according to their system usability rating.

The findings of the study respond to the demands of standardisation in a human-centric approach to manage handover and takeover between the vehicle and the human for SAE level 3 automation [1]. Besides the required compliance with the existing standards, the monitoring of the human state during the interaction with the AD functions also provides precious feedback on the machine's performance which can be used for improving the intelligent machine itself [19] and provides the foundations for further human-centred development with the flair of a humanistic AI control. Considering that ensuring dependability of such systems relies on AI approaches, it is still an open issue that lacks standard industrialisation solutions [15]. Ensuring dependability, despite the complexity and changing nature of the systems due to adaptation and learning, is a precondition for public trust and acceptance [58]. The findings of the study help to quantify human behaviour and define measurable parameters that are also working towards standardisation of the concept of safety-critical autonomous or AI-enhanced applications.

In addition, the presented work can support the establishment of trust and acceptance measures of AD in general and AI-based approaches in particular. Such measures may provide the foundation of necessary acceptance and standardisation related to human perception and trust in AD systems.
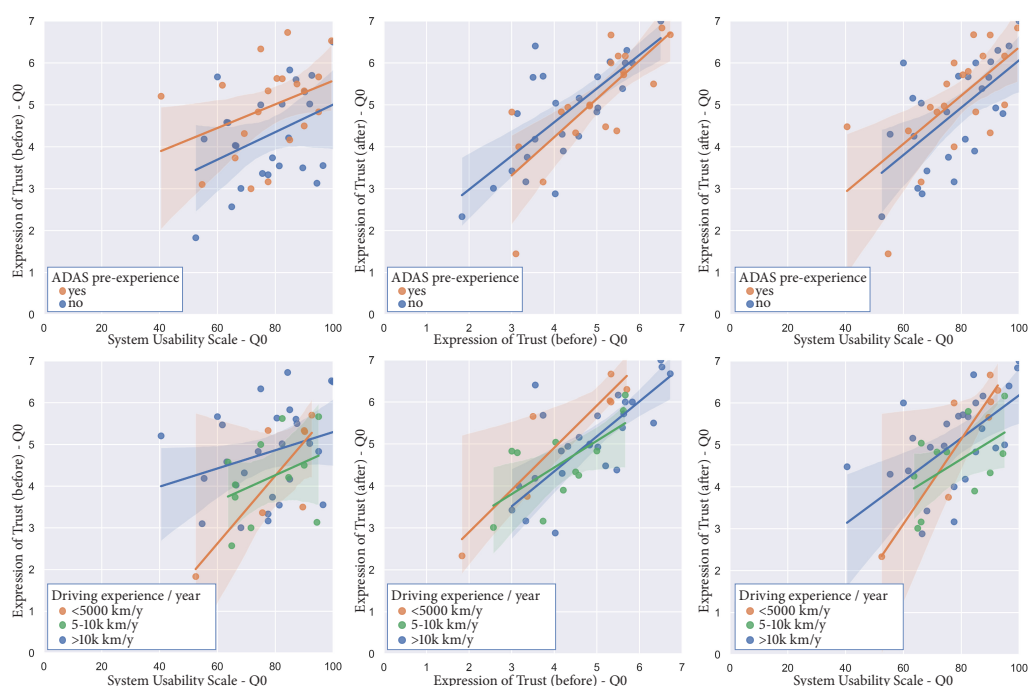
Figure 11. Pearson correlations with descriptive analysis of potential clusters regarding ADAS pre-experience and driving experience.

Limitations

A limiting factor in this study may be seen in its virtual reality nature within a simulator setting. This may provide an impression and sensation of safety which would not be experienced in reality and result in a more comfortable and less critical driver state in the simulator. We suggest that the direction of the effect is the same though the effect size may be smaller.

We would expect the NASA TLX questionnaire to lead to a higher level of workload when facing the same situation in a vehicle in a real-world scenario like on the road or a test track. This influence is expected to be lower for the Expression of Trust questionnaire due to the research focus on the difference between pre- and post-evaluation. In both cases, the participants are aware of the virtual character in the simulator. Therefore, the participants' rating focuses on the complete simulator session rather than on a single scenario.

Moreover, we wish to point out that the number of participants was not completely balanced between all groups analysed in the calculations. This results in a small number of participants within some specific groups and may lead to a limited generalisation for specific results while not harming the big picture of the findings.

Measuring the effects of AD on human perception was concluded on a highly advanced system, which may lead to a limited generalisation regarding the behaviour of the system within a high situational variance in the real world.

5. Conclusions

While the driving simulator offers a chance to immerse in AD functionality, it comes with the limitation posed by the virtual environment. The missing risk of harm can support the participant to gain sufficient impressions of the AD behaviour and therefore significantly increase the participant's trust in the system. When comparing the analysis of participants' impressions before and after the simulator session, the conducted experiment shows that this increase of trust holds for all demographic subgroups on a similar level. For example, the mean increase in trust resulting from the described experience in the driving simulator is measured 3.8% for study participants with and 13.7% without previous ADAS experience. Concerning the gender diversion an increase of trust of 5.4% for the male and 14.5% for the

female participants was measured. Group analyses do not reveal significant differences. Hence, we suggest a reduced need for precise and equally balanced demographic groups. Single significant differences within the subgroups are only recognised when using single inquiry lines (specific aspects within the questionnaires).

According to the participants' subjective perception, relying on the system usability scale, the system was able to provide sufficient performance (77.49%) for the conducted evaluation in a reliable virtual environment, substantiated by a homogeneous distribution of the demographic group mean ratings on the questionnaire on a high level.

For SAE level 4 AD in a simulator, participants' subjective workload appears to be low on all singular aspects of the NASA TLX. This reflects the definition of the automated driving mode with no objective task for the driver in the operational design domain. The factors age and driving experience have a significant influence on the participant's workload. Further, the measurement provided evidence that there is a significant relationship between the trust in the system and the participant's workload. In essence, the collected experimental data consistently indicates that the desired increase of trust (as a proxy to improve user acceptance) could be supported by reducing the need for user workload in vehicles.

At present, the integration of the highly automated and autonomous operation of safety-critical systems is still in an early stage. Therefore, an industry-agreed state of the art for measuring trust in autonomous or AI-based systems does not yet exist. Standards addressing the specific aspects of autonomous and/or AI-based technology for safety-critical systems are under development, providing essential further steps in such directions. Although this paper is not directly focusing on standardisation, the presented measurements addressing the establishment of trust and definition of measurable parameters create a foundation for further support for designers and engineers in the conception of safety-critical autonomous or AI-enhanced applications.

Future investigations into the complex combination of human drivers in an AD personal transportation system are required, using statistical analysis in combination with AI to understand and improve the future user trust in the AD technology.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Automated Driving
ADAS	Advanced Driver Assistant Systems
AI	Artificial Intelligence
ANOVA	Analysis of Variance
DDT	Dynamic Driving Task
EOT	Expression of Trust
GDPR	General Data Protection Regulation
GSR	General Safety Regulation
HMI	Human–machine Interface
HR	Heart rate
MANOVA	Multifactorial Analysis of Variance
MDPI	Multidisciplinary Digital Publishing Institute
NASA TLX	NASA Task Load Index
SAE	Society of Automotive Engineers
SSQT	Sequence Specific Questionnaire on Trust
SUS	System Usability Scale

References

1. On-Road Automated Driving (ORAD) Committee. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles*; SAE International: Warrendale, PA, USA, 2021. [CrossRef]
2. Davoli, L.; Martalò, M.; Cilfone, A.; Belli, L.; Ferrari, G.; Presta, R.; Montanari, R.; Mengoni, M.; Giraldi, L.; Amparore, E.G.; et al. On Driver Behavior Recognition for Increased Safety: A Roadmap. *Safety* **2020**, *6*, 55. [CrossRef]
3. Gasser, T.M.; Seeck, A.; Smith, B.W. Rahmenbedingungen für die Fahrerassistenzentwicklung. In *Handbuch Fahrerassistenzsysteme*; Springer: Berlin/Heidelberg, Germany, 2015. [CrossRef]
4. Marcano, M.; Tango, F.; Sarabia, J.; Castellano, A.; Pérez, J.; Irigoyen, E.; Díaz, S. From the Concept of Being "the Boss" to the Idea of Being "a Team": The Adaptive Co-Pilot as the Enabler for a New Cooperative Framework. *Appl. Sci.* **2021**, *11*, 6950. [CrossRef]
5. Jeon, M.; Walker, B.N.; Yim, J.B. Effects of specific emotions on subjective judgment, driving performance, and perceived workload. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *24*, 197–209. [CrossRef]
6. Egerstedt, M.; Hu, X.; Stotsky, A. Control of a car-like robot using a virtual vehicle approach. In Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171), Tampa, FL, USA, 16–18 December 1998; Volume 2, pp. 1502–1507. [CrossRef]
7. Deter, D.; Wang, C.; Cook, A.; Perry, N.K. Simulating the Autonomous Future: A Look at Virtual Vehicle Environments and How to Validate Simulation Using Public Data Sets. *IEEE Signal Process. Mag.* **2021**, *38*, 111–121. [CrossRef]
8. Wynne, R.A.; Beanland, V.; Salmon, P.M. Systematic review of driving simulator validation studies. *Saf. Sci.* **2019**, *117*, 138–151. [CrossRef]
9. Ogitsu, T.; Mizoguchi, H. A study on driver training on advanced driver assistance systems by using a driving simulator. In Proceedings of the 2015 International Conference on Connected Vehicles and Expo (ICCVE), Shenzhen, China, 19–23 October 2015; pp. 352–353. [CrossRef]
10. Saetren, G.B.; Pedersen, P.A.; Robertsen, R.; Haukeberg, P.; Rasmussen, M.; Lindheim, C. Simulator training in driver education-potential gains and challenges. In *Safety and Reliability-Safe Societies in a Changing World*; Taylor & Francis Group: Abingdon, UK, 2018.
11. FFG. Integrating Mental Training Techniques into Simulator Training. Available online: https://projekte.ffg.at/projekt/2758303 (accessed on 30 November 2021).
12. Schwarz, M.; Harfmann, J. Essential Tools for Safety Culture Development in Air Traffic Management. In *Aviation Psychology. Applied Methods and Techniques*; Koglbauer, I.V., Biede-Straussberger, S., Eds.; Hogrefe Publishing: Goettingen, Germany, 2021; pp. 33–50.
13. Talker, C.M.; Kallus, K. Anticipatorily Controlled Top-Down Processes Influence the Impact of Coriolis Effects. In Proceedings of the 18th International Symposium on Aviation Psychology, Dayton, OH, USA, 4–7 May 2015.
14. Azevedo Sá, H.; Jayaraman, S.; Esterwood, C.; Yang, X.J.; Robert, L.; Tilbury, D. Real-Time Estimation of Drivers' Trust in Automated Driving Systems. *SSRN Electron. J.* **2020**. [CrossRef]
15. Macher, G.; Akarmazyan, S.; Armengaud, E.; Bacciu, D.; Calandra, C.; Danzinger, H.; Dazzi, P.; Davalas, C.; Gennaro, M.C.D.; Dimitriou, A.; et al. Dependable Integration Concepts for Human-Centric AI-Based Systems. In *International Conference on Computer Safety, Reliability, and Security*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 11–23.
16. Veledar, O. New Business Models to Realise Benefits of the IoT Technology within the Automotive Industry. Ph.D. Thesis, Vienna University of Economics and Business—Executive Academy, Vienna, Austria, 2019. [CrossRef]

17. Bierzynski, K.; Calvo Alonso, D.; Gandhi, K.; Lehment, N.; Mayer, D.; Nackaerts, A.; Neul, R.; Peischl, B.; Rix, N.; Röhm, H. AI at the Edge, 2021 EPoSS White Paper. 2021. Available online: https://cora.ucc.ie/handle/10468/11495 (accessed on 29 December 2021).

18. Aptiv; Audi; Baidu; BMW; Continental; Daimler; FCA; HERE; Infineon; Intel; et al. Safety First for Automated Driving. 2019. Available online: https://www.aptiv.com/en/newsroom/article/automotive-and-mobility-industry-leaders-publish-first-of-its-kind-framework-for-safe-automated-driving-systems (accessed on 29 November 2021).

19. Bacciu, D.; Akarmazyan, S.; Armengaud, E.; Bacco, M.; Bravos, G.; Calandra, C.; Carlini, E.; Carta, A.; Cassarà, P.; Coppola, M.; et al. TEACHING-Trustworthy autonomous cyber-physical applications through human-centred intelligence. In Proceedings of the 2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS), Barcelona, Spain, 23–25 August 2021; pp. 1–6. [CrossRef]

20. ISO-International Standardization Organisation. ISO 26262 Road Vehicles-Functional Safety. 2018. Available online: https://www.iso.org/standard/68383.html (accessed on 30 November 2021).

21. ISO-International Organization for Standardization. ISO/WD PAS 21448 Road Vehicles-Safety of the Intended Functionality, Work-in-progress. Available online: https://www.iso.org/standard/70939.html (accessed on 20 August 2021).

22. Körber, M.; Baseler, E.; Bengler, K. Introduction matters: Manipulating trust in automation and reliance in automated driving. *Appl. Ergon.* **2018**, *66*, 18–31. [CrossRef]

23. Ward, C.; Raue, M.; Lee, C.; D'Ambrosio, L.; Coughlin, J.F. Acceptance of automated driving across generations: The role of risk and benefit perception, knowledge, and trust. In *International Conference on Human-Computer Interaction*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 254–266.

24. Dimitrakopoulos, G.; Uden, L.; Varlamis, I. Chapter 7-User acceptance and ethics of ITS. In *The Future of Intelligent Transport Systems*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 85–91. [CrossRef]

25. Zhang, T.; Tao, D.; Qu, X.; Zhang, X.; Zeng, J.; Zhu, H.; Zhu, H. Automated vehicle acceptance in China: Social influence and initial trust are key determinants. *Transp. Res. Part C Emerg. Technol.* **2020**, *112*, 220–233. [CrossRef]

26. Kaur, K.; Rampersad, G. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *J. Eng. Technol. Manag.* **2018**, *48*, 87–96. [CrossRef]

27. Kyriakidis, M.; de Winter, J.C.; Stanton, N.; Bellet, T.; van Arem, B.; Brookhuis, K.; Martens, M.H.; Bengler, K.; Andersson, J.; Merat, N.; et al. A human factors perspective on automated driving. *Theor. Issues Ergon. Sci.* **2019**, *20*, 223–249. [CrossRef]

28. Druml, N.; Ryabokon, A.; Schorn, R.; Koszescha, J.; Ozols, K.; Levinskis, A.; Novickis, R.; Nigussie, E.; Isoaho, J.; Solmaz, S.; et al. Programmable Systems for Intelligence in Automobiles (PRYSTINE): Final results after Year 3. In Proceedings of the 2021 24th Euromicro Conference on Digital System Design (DSD), Palermo, Spain, 1–3 September 2021; pp. 268–277. [CrossRef]

29. Klöckner, C.A. The dynamics of purchasing an electric vehicle–A prospective longitudinal study of the decision-making process. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *24*, 103–116. [CrossRef]

30. Ribeiro, M.A.; Gursoy, D.; Chi, O.H. Customer Acceptance of Autonomous Vehicles in Travel and Tourism. *J. Travel Res.* **2021**. [CrossRef]

31. Barth, M.; Jugert, P.; Fritsche, I. Still underdetected–Social norms and collective efficacy predict the acceptance of electric vehicles in Germany. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *37*, 64–77. [CrossRef]

32. Nordhoff, S.; Van Arem, B.; Happee, R. Conceptual model to explain, predict, and improve user acceptance of driverless podlike vehicles. *Transp. Res. Rec.* **2016**, *2602*, 60–67. [CrossRef]

33. Beggiato, M.; Krems, J.F. The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transp. Res. Part F Traffic Psychol. Behav.* **2013**, *18*, 47–57. [CrossRef]

34. Macher, G.; Druml, N.; Veledar, O.; Reckenzaun, J. Safety and Security Aspects of Fail-Operational Urban Surround perceptION (FUSION). In *Model-Based Safety and Assessment*; Papadopoulos, Y., Aslansefat, K., Katsaros, P., Bozzano, M., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 286–300. [CrossRef]

35. Hartwich, F.; Beggiato, M.; Krems, J.F. Driving comfort, enjoyment and acceptance of automated driving–effects of drivers' age and driving style familiarity. *Ergonomics* **2018**, *61*, 1017–1032. [CrossRef]

36. Hanzl, G.; Haberl, M.; Eichberger, A.; Fellendorf, M. Human Driver's Acceptance of Automated Driving Systems Based on a Driving Simulator Study. In *International Forum on Advanced Microsystems for Automotive Applications*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 186–195.

37. Watzenig, D.; Horn, M. Introduction to Automated Driving. In *Automated Driving: Safer and More Efficient Future Driving*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 3–16. [CrossRef]

38. Veledar, O.; Armengaud, E.; Happ Botler, L.; Damjanovic-Behrendt, V.; Jaksic, S. Steering Drivers of Change: Maximising Benefits of Trustworthy IoT. In *Systems, Software and Services Process Improvement. EuroSPI 2021. Communications in Computer and Information Science*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 1442, pp. 663–674.

39. Jian, J.Y.; Bisantz, A.M.; Drury, C.G. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Int. J. Cogn. Ergon.* **2000**, *4*, 53–71. [CrossRef]

40. U.S. General Services Administration—Technology Transformation Services. System Usability Scale (SUS). 2013. Available online: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html (accessed on 12 January 2022)

41. Said, S.; Gozdzik, M.; Roche, T.R.; Braun, J.; Rössler, J.; Kaserer, A.; Spahn, D.R.; Nöthiger, C.B.; Tscholl, D.W. Validation of the Raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) Questionnaire to Assess Perceived Workload in Patient Monitoring Tasks: Pooled Analysis Study Using Mixed Models. *J. Med. Internet Res.* **2020**, *22*, e19472. [CrossRef] [PubMed]

42. Grier, R.A. How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2015**, *59*, 1727–1731. [CrossRef]

43. UNESCO Institute for Statistics. International standard classification of education: ISCED 2011. *Comp. Soc. Res.* **2012**, *30*. [CrossRef]

44. Clement, P.; Danzinger, H.; Veledar, O.; Koenczoel, C.; Macher, G.; Eichberger, A. Measuring trust in automated driving using a multi-level approach to human factors. In Proceedings of the 24th Euromicro Conference on Digital System Design (DSD), Palermo, Italy, 1–3 September 2021; Leporati, S.F., Vitabile, A.S., Eds.; IEEE Computer Society: Palermo, IT, USA, 2021; pp. 410–417. [CrossRef]

45. Helldin, T.; Falkman, G.; Riveiro, M.; Davidsson, S. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Leeds, UK, 28–30 October 2013*; Terken, J., Ed.; ACM: New York, NY, USA, 2013; pp. 210–217. [CrossRef]

46. Bangor, A.; Kortum, P.; Miller, J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud.* **2009**, *4*, 114–123.

47. Grier, R.A.; Bangor, A.; Kortum, P.; Peres, S.C. The System Usability Scale. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2013**, *57*, 187–191. [CrossRef]

48. Brooke, J. SUS: A quickdirty usability scale. In *Usability Evaluation in Industry*; Taylor and Francis: London, UK, 1996; p. 252.

49. Rauer, M. Quantitative Usablility-Analysen mit der System Usability Scale (SUS)-Nachrichten, Tipps & Anleitungen für Agile, Entwicklung, Atlassian-Software (JIRA, Confluence, Bitbucket, ...) und Google Cloud. 2011. Available online: https://blog.seibert-media.net/blog/2011/04/11/usablility-analysen-system-usability-scale-sus/ (accessed on 20 April 2021).

50. Index, The NASA TLX Tool: Task Load. TLX @ NASA Ames-Home. Available online: https://human-factors.arc.nasa.gov/groups/TLX/ (accessed on 19 April 2021).

51. EU. General Data Protection Regulation EU3016/0679: GDPR. Available online: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679 (accessed on 19 April 2021).

52. CRUDEN. Automotive Driving Simulators-Cruden. Available online: https://www.cruden.com/automotive-driving-simulators/ (accessed on 29 December 2021).

53. VIRES. VTD Virtual Test Drive. Available online: https://www.mscsoftware.com/de/virtual-test-drive (accessed on 19 April 2021).

54. AVL List GmbH. AVL VSM™ Vehicle Simulation. Available online: https://www.avl.com/-/avl-vsm-4- (accessed on 19 April 2021).

55. AVL List GmbH. AVL Model.Connect™. Available online: https://www.avl.com/de/-/model-connect- (accessed on 19 April 2021).

56. Clement, P.; Danzinger, H.; Quinz, P.; Hillbrand, B.; Hartavi, A.E.; Kasikci, K.Z. Assessment Concept for TrustVehicles. In *Enhanced Trustworthiness and End User Acceptance of Conditionally Automated Vehicles in the Transition Period*; Watzenig, D., Schicker, L.M., Eds.; Lecture Notes in Intelligent Transportation and Infrastructure; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 107–128. [CrossRef]

57. TrustVehicle H2020 Project. Experiment Scenarios Video Description of Scnearios. Available online: https://www.youtube.com/channel/UCdsw-Md5OxfsgFyQY7DtXpA (accessed on 4 November 2021).

58. Macher, G.; Diwold, K.; Veledar, O.; Armengaud, E.; Römer, K. The Quest for Infrastructures and Engineering Methods Enabling Highly Dynamic Autonomous Systems. In *Systems, Software and Services Process Improvement*; Walker, A., O'Connor, R.V., Messnarz, R., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 15–27.