Measuring trust in automated driving using a multi-level approach to human factors*

Philipp Clement
ADAS/AD Dev. and Validation
AVL List GmbH
Graz, Austria
philipp.clement@avl.com

Herbert Danzinger
Driving Attributes Assessment and Dev.
AVL List GmbH
Graz, Austria
herbert.danzinger@avl.com

Omar Veledar
Research and Techn. Development
AVL List GmbH
Graz, Austria
omar.veledar@avl.com

Clemens Könczöl
Institute of Psychology
University of Graz
Graz, Austria
clemens.koenczoel@uni-graz.at

Georg Macher
Institute for Technical Informatics
Graz University of Technology
Graz, Austria
georg.macher@tugraz.at

Arno Eichberger
Institute of Automotive Engineering
Graz University of Technology
Graz, Austria
arno.eichberger@tugraz.at

Abstract—As the driving is shifting towards automation, the maximization of related benefits would profit from improved user acceptance of the new technology. Studies suggest a strong connection between acceptance and trust in technical solutions. We investigate the improvement of user trust to driving automation through demonstrations that carried on a sophisticated driving simulator. The study correlates subjective data with objective psycho-physiological measurements. The multi-factorial and multivariate analysis of variance investigates the influence of learning effects and pre-experience with ADAS on trust. Results show improvement in trust through user interaction with a human-machine interface of the demonstrated AD system, hence illustrating the relevance of human-centered development processes. The conclusion is supported by the observation of driver cardiac signals.

Index Terms—In cabin measurement, HMI, trust, comfort, driver state, ADAS/AD, multi-level approach, psychophysiology, heart-rate

I. INTRODUCTION

Autonomous Driving (AD) is edging towards an inevitable reality. The technological progress in combination with market expectations are requesting high research needs in terms of human machine interaction. [18] The key contributor to future road safety is the promise of reduced number and the consequences of road accidents. The majority of those are caused by humans [35], whose reaction times, distraction levels, tiredness or influence of (legal/illegal) performance altering substances, where those deficiencies could be prevented by automated vehicle controls. The complementary AD features contribute to improved traffic flow and increased road capacity, include applications, such as intelligent route-planning or platooning [36]. However, the positive benefits are matched by a possible lack of acceptance of the new technology [33], which is related to cyber-security [26] and privacy [2]. The reduced acceptance level is also correlated to the lack of trust in technical solutions [10] and safety [27].

Demonstrations with sophisticated simulators could improve trust levels by revealing the real benefits of AD to the human user. Such virtual environment studies yield insights into user behaviour in safety-critical scenarios without compromising user safety. The objective data on soft factors (trust, comfort and drowsiness) are gathered in a series of scenarios. Three major drivers of the development are legislation [32], safety and comfort [25].

Besides, standardisation demands a human-centric approach to manage handover and takeover between the vehicle and the humans for SAE Level 2 and 3 automation [30]. The same complies with the General Safety Regulation (GSR) [12]. A level of driver interaction, including exchange about the state of the vehicle and monitoring of the driver state e.g., the Driver Drowsiness and Attention Warning (DDAW) is essential when the driver must take over the control of the vehicle, for the environment monitoring task (SAE L2) or the availability as fallback level (SAE L3). Vehicles must also fulfil safety requirements, which are vital for the end-users. Additional demands come from Euro NCAP in terms of Safety Assist test procedure [13], which centres around occupant monitoring. They focus on the minimisation of human errors through occupant monitoring and warning, as some studies suggest that 90% of the accident are caused by human mistakes (purposeful or erroneous) [13]. Comfort, which is difficult to assess objectively, is another crucial quality for the users [10]. Comfort distinguishes different service providers and can also contribute to the perceived safety, which is either a selling factor or a blocker.

Driving simulator studies support the full development cycle from the concept, through the implementation, to the validation and verification phase. Their unique value is safe

testing of safety-critical situations (e.g., drivers falling asleep). Their major benefit is repeatable testing with exactly the same variables and covariates for a given scenario at the concept, design, validation and verification phase. This eases monitoring and development. The high reliability enables assessment of the forthcoming events in the development and adoption of goals in agile structures. In on-road testing it is impossible to reach the same real driving repeatability due to the dynamic environment and road conditions. That leads to a high internal validity of the testing method.

Simulator benefits from cost efficiency, internal and validity, precise control of co-variate, independent repeatability, high situational specificity, exposure to critical situations, safe environment, etc. This work exploits the offered benefits to examine trust and end-user acceptance of AD. If defining trust as a belief in the dependability of the AD system, its strong link to acceptance and use of automated technology [28] is crucial for the impact of AD. This paper investigates the improvement of trust through pre-existing and ongoing human-machine interaction with an AD system.

II. METHOD

This section describes the setting and assets employed in the study. The process executes a tailored set of simulator scenarios and investigates the prospect of growing trust in AD.

A. Sample

The experiment was completed by 55 of 60 registered subjects. Upon removing 4 incomplete data sets and 2 sets with outliers including inconsistent answers between control questions, the number of complete data sets and valid results for the analysis is 49. The subjects are grouped to evaluate demographic differences and group interactions with other factors. The groups are: gender (28 male, 21 female, 0 neutral), age (21 between 18-29 years, 20 between 30-45 years and 8 between 45-65 years), yearly driving range (9 below 5000 km/y, 12 between 5001-10000 km/y, and 29 over 10001 km/y), driving assistant experience (23 without, and 26 with) and educational level (37 higher than Bachelor's or equivalent degree and 12 below according to ISCED) [14].

B. Scenarios

The scenario selection follows the method of assessment concepts for TrustVehicles [9]. Ten scenarios with one or more sequences are chosen with changes in the driving mode between them. Two controller modes are applied: a comfortable mode with moderate deceleration and larger gaps and a sporty mode with higher deceleration and smaller gaps. The scenarios contain take over and function decrease due to dirt (scenario 1 & 2), follow constant and function decrease due to weather change (scenario 3 & 4), emergency brake (scenario 5, 6, & 7), and lane change situations and weather change (scenario 8, 9, & 10). All scenarios are in fully AD mode, so each subject experiences the same situation. This is a requirement for a high internal validity and precise control of covariates. A

representative scenario (emergency brake) is described below. Matching sequences are clustered for the evaluation.

The emergency brake scenario is applied in a narrow one way street with vehicles parked on both sides of the moving ego-vehicle. Upon approaching this location the ego-vehicle slows down from 30 km/h to an appropriate velocity for the road type (15 km/h). The deceleration is mandated by the limited perception boundaries due to the reduced view of sight of the vehicle sensors in such an environment. The ego-vehicle continues at a constant velocity until it detects a child attempting to cross the road at a location with no labelled road crossings, between the parked vehicles from the right side (see figure 1). The sudden appearance prompts a swift reaction of the vehicle controls. The breaks are activated. The key parameter in this scenario is the timing and the intensity of the breaking. The scenario cluster is based on the following order:

S5 Break early (comfortable controller configuration)
S6 Late breaking (sporty controller configuration)
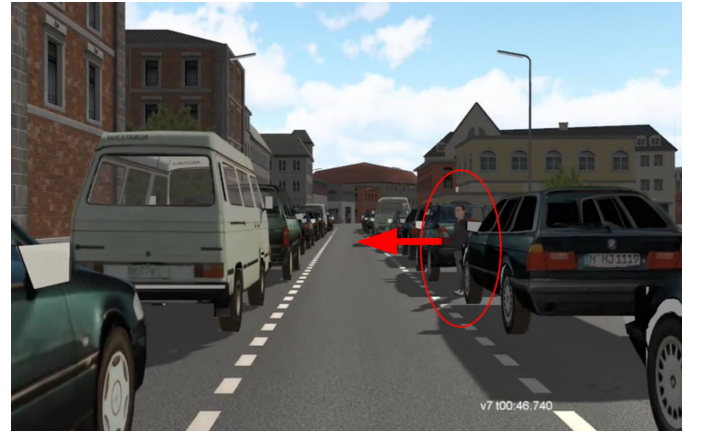S7 Early breaking again (identical to the first scenario)

Fig. 1: Visualisation: child appears between parked vehicles and moving direction (red arrow) of scenario 6.

The fifth scenario is used to monitor driver reactions to a comfortable response to the child crossing the road. The resulting measurements are treated as the baseline measurement. The sixth scenario intends to measure the drivers reaction to a sporty controller reaction, thus creating a more critical situation. The intention is to convey to the subject that even in the event of the late braking, the situation is safe. The seventh scenario is compared to the previous two scenarios to observe the impact of the learning effect on the trust in the technology.

C. Study materials

The study engages a multilevel approach to subjective and objective data, as well as psycho-physiological parameters to examine the influence of the AD system on trust. The created "Sequence-Specific Questionnaire on Trust" (SSQT) results from the experience of driving behaviour testing and development [3]. Its three key questions focus on comfort (Q1), trust (Q2) and overall perception (Q3). It also includes

sequence-specific questions. The rating scale ranges from 1 (*very bad*) to 10 (*very good*) or in some sequence-specific questions from 1 (*bad* e.g., *too early*), through 10 (*good* e.g., *just right*) to 20 (*bad* e.g., *too late*). The three standard questions for all sequences are:

Q1 How comfortable did you feel?
Q2 How much confidence did you have in the vehicle?
Q3 How was the overall perception?

Q1 How comfortable did you feel?

not at all   ok   very comfortable

Fig. 2: Cutout sample of the center console questionnaire (Q1).

The three standard questions are complemented by the sequence-related questions, which assess the behaviour of the AD controller and its perceived performance. For scenarios 5 to 7, these questions are *How do you evaluate the reaction time of the vehicle until the first reaction in front of the pedestrian? How do you evaluate the distance from the stopped vehicle to the crossing pedestrian?* and *How do you evaluate the time until the vehicle starts again after the pedestrian has left the roadway?.* The SSQT accompanied by a short video of the scene to aid subjects' memory. The answers are entered on a slider scale (color, text and smiley coded) shown in figure 2. Trust is related to question number two, and supported by question one (comfort) and question three (overall perception). These questions indicate subjects' experience of the situation. The detailed experience is explored in the sequence-specific questions.

The heart rate is measured by POLAR H10 chest belt [29] and is communicated via ANT+ and Bluetooth to the co-simulation framework. The demographic data and subjective experience are quantified using questionnaires prior, during and on completion of the test drive.

Additional assessment materials are used but are not evaluated for this paper. NASA TLX [23] measures subjective assessment of mental demand, physical demand, temporal demand, performance, effort and frustration, during the post-test phase. The expression of trust questionnaire is a variation of [21], which is built on the original trust in automation scale [24]. It enquires the trust in AD functions using a seven-point Likert scale from *do not agree at all* to *agree completely*. Questions are aimed at answering if *the trust in AD can be raised through a safe experience*. The usability of the system under test is assessed using usability scale [1] with a 10 question questionnaire slightly adapted to the needs of the study by following the actual advice [20], also during the post-test phase. The eye tracking is implemented using a Pupil Labs Core device [22] to monitor the gaze direction, pupil size and blink rate. The connection to the co-simulation framework uses a USB interface to handle the video data volume.

D. Environment and procedure

The controlled laboratory environment study engages 60 volunteers of diverse ages and nearly equal gender division recruited over official websites and inter company information. These are licensed drivers with no driving simulation experience. They follow an identical set of procedures and scenarios using the same set of hardware and software tools. Test sessions start with an familiarization phase of 10 minutes, which is followed by a set of AD scenarios including driving in rural, city, country, narrow parking, construction site and other environments.

E. Equipment and Techniques

The technical setting consists of the simulator itself, the background hardware (e.g., computers, sensors) and the software combining the hardware into the test environment. The simulator is a moving hexapod platform with a real vehicle cut out containing the driver seat and the complete driver environment (the cockpit). The virtual environment is visualised by projection onto a 180° canvas by 3 video projectors with a high frame rate and resolution (100Hz, 4K) for realistic simulation and reduction of motion sickness. Additional rearview mirror screens increase the field of view and improve realism. The motion system, characterised by six degrees of freedom, offers many manoeuvring possibilities, as required for simulation of the following motion types: three dimension translational and pivoting around each axis. The simulator acceleration is limited to 6m/s² with a manoeuvring possibility of approximately 1m in all three dimensions (x/y/z) and a rotational capability. In terms of the reduced motion sickness experienced by the subjects (section II-A), the improvement is traced back to an optimized motion cueing and adopted platform motion to real vehicle test (external validity). The platform is limited to perform real road behaviour by its technical boundaries. Biometric data is recorded with a range
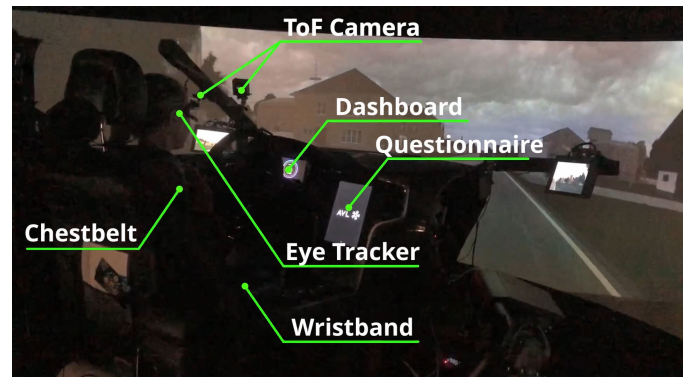

Fig. 3: Visualisation of the cockpit with marked sensors and center console.

of sensors during tests within the cockpit: chest belt for the heart rate (HR) and the heart rate variability (HRV), Time of Flight (ToF) cameras, eye tracker and eye gaze monitor, head and body movement monitors partly visualised in figure 3. The sensor combination differs subject to the study, topic,

and availability of the evolving sensors. The driver monitoring sensors enable timely and appropriate driver alerting actions about an impeding driver takeover of vehicle controls, as specified for the L3 automated vehicles. Additional subjective data is gathered through pre-simulation, interim and post-simulation questionnaires.

A co-simulation framework with integrated AD functions integrates software and hardware to facilitate the usability of all components. The framework also enables simulation of scenarios that are unlikely to be tested in real driving situations due to safety concerns. Two main technical challenges associated with the setup of the driving simulator are posed by the data time synchronisation and synchronisation of simulator movements. The data synchronisation is tackled by component interconnection using the AVL Model.CONNECT™ [4], a co-simulation platform that enables data acquisition and monitoring from a single communication centre. The advantage of a centralised data collection point is the chance to mitigate and avoid the realisation of unforeseen simulation interruption or data loss through a complete run-time overview of the system. Data with divergent sampling rates are partially handled by first and second-order interpolation and extrapolation techniques in combination with the Nearly Energy Preserving Coupling Element (NEPCE) method [6], which minimises the coupling errors and increases numerical stability.

The motion platform is controlled by AVL VSM™ [5], which simulates the vehicle behaviour and translates expected accelerations into precise simulator motions. This is essential for a realistic feeling and reduction of motion sickness. VIRES VTD™ [34] supports road visualisation by offering environment setup including road networks, traffic infrastructure and other road users. It also aids the data distribution from simulated perception sensors, as needed for influencing vehicle control outputs. The structured data distribution and synchronisation of the sensor outputs are crucial for the optimisation of the simulator and control strategy. The simulator also offers freedom in terms of hardware integration. That enables complex manoeuvring of the modular physical platform. The environment enables usage of bulky hardware that can not be installed in vehicle testing due to limitations in terms of physical space or technical integration, or safety.

While other experiments [16] [15] [17] are based on the same methodology, the unique character of this study is based on a specific combination of the physiological measurement data with the subjective purpose-designed questionnaire within a study that targets user trust.

F. The Experimental Process

Figure 4 depicts the process a single subject is exposed to. Through upfront information, a subject arriving at the test site is already familiar with the testing procedure and data to be collected. They are going through six major test phases.

A subject arriving at the test site is familiarised with the information and the agreement sheets at their own pace at the start of phase one. They are offered a verbal question and answer session, which also includes a mandatory introduction

of the relevant General Data Protection Regulation (GDPR) [11] aspects. It is crucial at this stage that the subject can stop the testing at any time and withdraw from the study. A pre-testing questionnaire, the expression of trust, is provided to the subject. This questionnaire is answered twice, once in phase one and once in phase six.

The second phase equips the subject with chosen sensors. The subject is introduced to the cockpit environment i.e. steering, braking, human-machine interfaces (e.g., indicators, dashboards) and the automated subjective feedback system in the centre console. Example questionnaires are hosted to familiarise the subject with the process and reduce the influence of an unknown system in the later answers.

The third (testing) phase lets the subject experiment with the behaviour of the moving platform for about 10 minutes. They are also confronted with several test questions. All sensors and questions are processed to confirm proper behaviour. The starting sub-phase does not include platform activation to allow gradual familiarisation of the subjects. The second sub-phase triggers the motion platform.

The fourth test phase drives automatically through several scenarios. Every scenario completion causes freezes of the environment to raise associated SSQTs. On completion of the SSQT, the subsequent scenario follows. The process is repeated for all scheduled scenarios and takes around one hour. During phases 3 and 4, all sensor data is recorded.

In the fifth phase, the subject leaves the cockpit and gets all the sensors and equipment removed. They get involved in a talk to check their subjects mental and physical health.

The subjects answer further questions in the last phase. The procedure keeps the subjects occupied and under observation for at least 30 minutes after the test. This prevents post-test motion sickness outside the controlled test environment. The questionnaires at this phase are the NASA TLX test [23], the expression of trust (again), the system usability scale [1] and a demographic questionnaire to relate their results to demographic groups. Finally, optional free form feedback enables reporting likes or dislikes of the complete participation.

III. RESULTS

The repeated measurement data is subjected to multifactorial and multivariate analysis of variance. The calculations are aimed at investigating differences between the scenarios, the driving mode, the driving experience and the ADAS experience. A fraction of the results is presented in this section, aiming to establish the link and likely trust increase through user interaction with an AD system.

A. Questionnaires

The results are focused on example scenarios 5.1, 6.1 and 7.1 (described in section II-B). For each SSQT question, we calculate a single analysis of variance with repeated measurements on three points of time. Significant differences emerge over time (p_t) within each question (see table I), interaction (p_{int}), and group (p_{gr}). The significant differences between
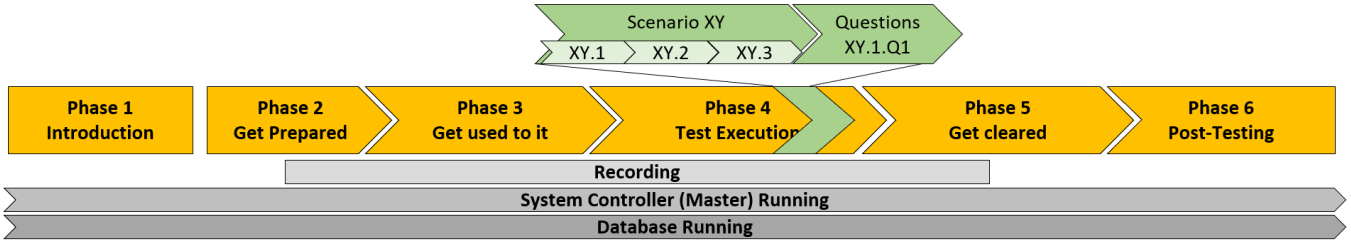
Fig. 4: The method of processing one subject throughout the complete test

different times of measurements are ($F=4.571$, $p_t=0.013$, $\eta^2=0.087$), ($F=3.112$, $p_t=0.049$, $\eta^2=0.061$) and ($F=3.121$, $p_t=0.49$, $\eta^2=0.061$) for questions 1, 2 and 3 respectively. The outcome is depicted in figure 5.
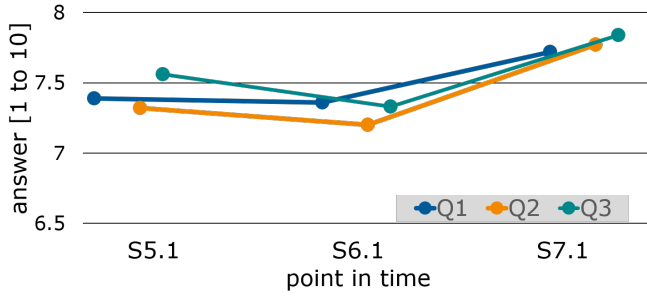

Fig. 5: Subjective ratings of comfort (Q1), trust (Q2) and overall perception (Q3) in scenarios 5, 6, and 7

Figure 5 shows the changes over time regarding the mean values of comfort (Q1), trust (Q2) and overall perception (Q3) in the example scenarios (5.1, 6.1 and 7.1). While comfort decreases from scenario 5 (comfort driving mode) to scenario 6 (sports driving mode), an increase can be shown from scenario 6 to scenario 7 (comfort driving mode), while values in scenario 7 outrun the other scenarios. Results for trust and overall perception show similar development and characteristics between the different scenarios.

TABLE I: Values of Q1, 2, and 3 for scenarios 5, 6, and 7

	Scenario 5		Scenario 6		Scenario 7		p_t
	mean	sd	mean	sd	mean	sd	
Q1	7.39	2.03	7.36	2.38	7.72	2.12	0.013
Q2	7.32	2.11	7.20	2.24	7.77	2.17	0.049
Q3	7.56	1.93	7.33	2.22	7.84	2.05	0.049

Results regarding experience of ADAS functionality depending on the three SSQT questions are shown in table II and in the figures 6, 7 and 8. For each scenario-specific question, we calculated a single analysis of variance with repeated measurements on three points of time including the factor pre-experience on ADAS.

As shown in table II, results reveal a trend towards a significant interaction ($F=2.171$, $p_{int}=0.119$, $\eta^2=0.044$) from the group and time on question 3. Significant differences can be seen in the main effect time in question 1 ($F=4.521$, $p_t=0.013$, $\eta^2=0.088$) and 2 ($F=3.102$, $p_t=0.049$, $\eta^2=0.062$)

TABLE II: Values of Q1, 2, and 3 for scenarios 5, 6, and 7 and interaction with ADAS experience. Acronyms for p values are gr...group, t...time, int...interaction.

	AD Exp.	Scenario 5		Scenario 6		Scenario 7		p_*
		mean	sd	mean	sd	mean	sd	
Q1	yes	7.11	2.31	6.76	2.41	7.65	2.15	p_{gr} 0.485
	no	7.64	1.75	7.29	2.37	7.79	2.13	p_t 0.013 p_{int} 0.619
Q2	yes	7.12	2.30	7.10	2.43	7.82	2.35	p_{gr} 0.700
	no	7.50	1.95	7.49	2.09	7.72	2.04	p_t 0.049 p_{int} 0.430
Q3	yes	7.41	2.08	6.97	2.34	7.93	2.19	p_{gr} 0.617
	no	7.70	1.81	7.66	2.09	7.77	1.96	p_t 0.045 p_{int} 0.119

and moreover a significant difference regarding the main effect "ADAS pre-experience" ($F=3.198$, $p_t=0.045$, $\eta^2=0.064$) shows up. Details are depicted in the figures 6, 7 and 8 respectively.
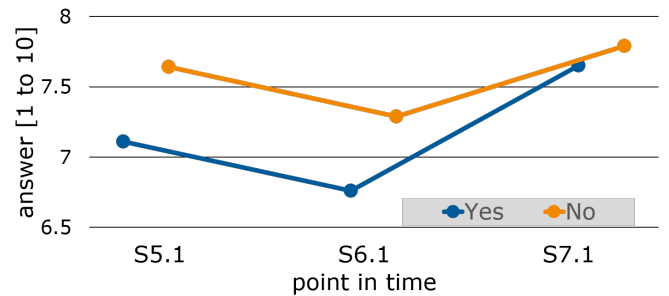

Fig. 6: Subjective ratings of comfort (Q1) with interaction to pre-existing ADAS experience in scenarios 5, 6, and 7.
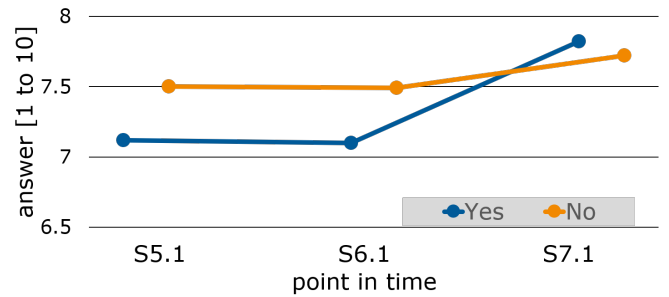

Fig. 7: Subjective ratings of trust (Q2) with interaction to pre-existing ADAS experience in scenarios 5, 6, and 7.
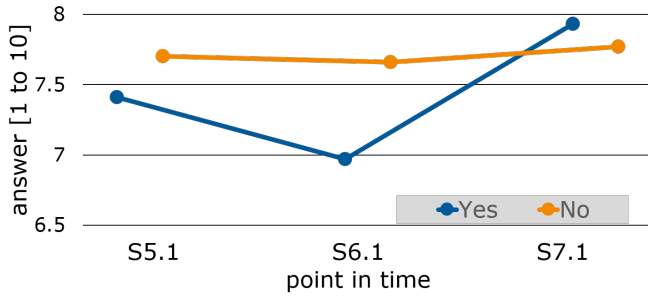
Fig. 8: Subjective ratings of overall perception (Q3) with interaction to pre-existing ADAS experience in scenarios 5, 6, and 7.
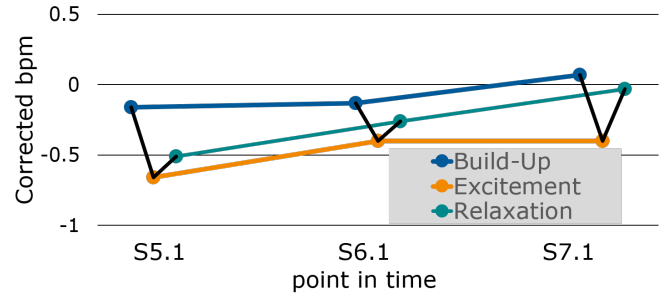

Fig. 9: baseline corrected & normalized heart rate values depending on the scenarios 5, 6, and 7 and the "Build Up", "Excitement", and "relaxation" intervals.

Figure 6 (Q1), figure 7 (Q2) and figure 8 (Q3) provide a similar picture regarding the progress of comfort, trust and overall perception over the three scenarios. Participants who already had an ADAS experience show lower ratings on the functionality at the first two emergency brake scenarios (S5.1 and S6.1) and show higher ratings in the third scenario (S7.1) for all three questions. They also show a lower rating on comfort (Q1) and overall perception (Q3) in the late braking scenario (S6.1) than the participants without ADAS experience.

B. Heart rate

Physiological parameters (e.g. the heart rate) were baseline corrected and normalized using an overall baseline regarding participants interindividual physical conditions. The correction was done for each subject per scenario to avoid cross interaction and effects of diverse subjects individual physical conditions during the progress in the experimental process. These values are compared to the intervals of interest of each subject. The intervals of interest were defined for the scenarios around the excitement actions with a predefined length. In scenario 5, this is the moment when the child appears between the parked vehicles visually for the vehicle sensors. Also, a build-up interval and a relaxation interval were defined to the same extent.

TABLE III: Values for HR-intervalls corrected and normalized for scenarios 5, 6, and 7

	Scenario 5		Scenario 6		Scenario 7	
	mean	sd	mean	sd	mean	sd
Build Up	-0.16	0.85	-0.13	0.99	0.07	0.93
Excitement	-0.66	0.72	-0.40	0.83	-0.40	0.92
Relaxation	-0.51	0.80	-0.26	0.67	-0.03	1.19
p_t	1.6e-6		0.048		0.044	

Table III shows means, sd and p values for the analysis of variance regarding the repeated measurements within scenario 5, 6 and 7. A highly significant difference is revealed within the points of measurement in scenario 5 ($F=17.441$, $p_t=0.000$, $\eta^2=0.401$) and significant differences in scenario 6 ($F=3.213$, $p_t=0.048$, $\eta^2=0.110$) and 7 ($F=3.334$, $p_t=0.044$, $\eta^2=0.118$).

Figure 9 shows heart rate reactivity in the "build up" interval being similar to the overall baseline, while the strongest reactions are in the excitement phase. In the relaxation interval, reactions in the heart rate are higher than in the excitement phase but still have not reached the level of the buildup interval, pictured as black line for each scenario. Moreover, fewer heart rate reactions can be seen in scenario 6 when compared to scenario 5, and in scenario 7 in relation to scenario 6. Scenario 7 heart rate in the relaxation interval has almost reached the level of the buildup phase.

IV. DISCUSSION

The evaluated results demonstrate that trust evolves with participants experience of the system in the present scenarios. Differences are seen in various levels of human reactions and behaviour. Even if starting from a relatively high trust level, increasing trust in AD is evident in all sequential experiments. This uncovers ongoing learning effects for the participants while they interact with an AD system. Results on this progressive gain of trust for the given scenarios are shown in figures 5 and 9.

Moreover, trust tends to differ regarding participants experience with AD ($p_{int}=0.119$). While participants who had no experience of AD tend to show constant overall perception, participants with pre-existing experience with AD show a decrease from scenario 5 to 6 and an increase from scenario 6 to 7 exceeding the rating of participants who had no experience with AD. Regarding those two groups, we notice the possible impact of pre-existing learning effects as well as learning effects that appear during the current human-machine interaction with ADAS systems.

Learning effects during the current human-machine interaction as well as pre-existing effects may have an important role in the development of trust in ADAS systems. In contradiction to other studies [28], which have not linked previous ADAS experience to trust, we report that ADAS experience might initially reduce trust in the system, but then it gains support for trust faster than with no ADAS experience. The reasons for the initial drop in trust remain unclear. The cause could be hidden in the subjects' negative experiences. Another hypothesis is that humans need sufficient experience and knowledge about the system functionality to gain trust. Both perspectives encourage the early learning effects of future

drivers while eliminating negative experiences through human-centered development of ADAS at the same time.

Cardiovascular reactions differ during the intervals (build up, excitement, relaxation) and between scenarios. While we identify strong heart rate reactions within the excitement interval, heart rate in the relaxation interval approximates to the baseline. The cardiac deceleration within the excitement interval in comparison to the build-up can be explained with participants orienting response within the situation, as the organism reacts to foster information input. That is consistent with literature [19] [8], which sees cardiac deceleration as a part of the freeze reaction in the orienting response, which helps humans to prepare for an adequate reaction by enhanced information acquisition, stimulated by the Amygdala. The cardiac deceleration might also refer to the behavioral inhibition system (BIS) which regards an inhibition of behavior by the hippocampus as the central structure, to avoid immediate actions if the system cannot predict the result of the planned action [7]. It then inhibits the behavior and evaluates the situation as a risk assessment by information processes. Observed cardiac acceleration in the relaxation interval refers to a regeneration of the cardiovascular system as the situation was, from drivers perspective, successfully coped by ADAS.

Differences between scenarios show less reactivity in the cardiovascular system within the excitement and relaxation intervals from scenario 5 to 7 in comparison to the build-up. This may be explained by current learning effects regarding the functionality and reliability of the system within the situation and therefore result in a gain in trust, than with a habituation process, as the presentation of three stimuli seems insufficient to trigger habituation processes. Results on cardiovascular reactivity, therefore, support the findings from the analysis on subjective ratings and supplement the evaluation with psycho-physiological concepts.

Results gained through simulator studies lack the real driving experience. The lack of possibility to experience unsafe or harmful events and from the safety perspective, the missing risk of injury might influence the intensity of participants reactions in the subjective ratings as well as in physical states while the direction of the results still correlates. Therefore simulator studies have an advantage in early-stage development, where the direction of the results is more important than their intensity. The differences in the intensity between the simulator and field results can be approached from different perspectives. These explanations varies to how well the simulator replicates the movements of a real vehicle [31]. Pre-testing evaluations of the simulator can contribute to their further development to increase external validity.

It seems to be crucial to allow the subjects to slip into the virtual reality and to avoid interactions from outside of this virtual world. To increase the validity of simulator studies participants must get used to the simulator and its behaviour in the pretesting phase. Surprisingly, many subjects reported in the free-text feedback that they felt more comfortable in the pretesting phase with activated platform movement than in the static phase. Data from this pretesting phase is excluded from the analysis. Another way to keep participants mentally in the simulation is to enable answering subjective questions within the cockpit (central dashboard) shortly after the situations and slip in and out of new scenarios as smooth as possible. That means that future studies might need to continue the simulation while subjects are answering the subjective questions. Once they are ready to continue with the testing, the simulation software should make a smooth transition back into the next scenario. Psycho-physiological methods contribute to this issue, as data acquisition happens right in the situation without interrupting the simulation or changing participants attention in the situation.

Considering figure 5, trust into the system remains stable, while the comfort and overall perception decrease slightly as expected by confronting the subjects with a more critical situation (later and more harsh braking in front of the child). Whereas in the following scenario (7) with repeated early braking situation, all three answers are raising over the first experience again. Trust raises significantly p_t=0.049 and comfort p_t=0.012 and overall perception p_t=0.048 follow this trend.

In summary, our finding is that users who gain information and experience with the system, its capabilities, and events (e.g., later breaking), provoke reactions with less intensity, as they rely on their learning about the system. However, when experiencing a critical event (i.e. later breaking) for the first time, with no knowledge about the system, it is reasonable to expect an intense driver response. Once the driver trust evolves through learning, we expect a generally less intense response to AD behaviour. This is confirmed, as we observe higher acceptance of critical situations due to the trust increase. These findings bring us a step closer towards answering our research question, which aims to establish a relationship between improvement of trust and ongoing user interaction with an AD system.

V. CONCLUSION

We explore the improvement of trust through ongoing user interaction with an AD system. The study merges responses to sequence-specific questions and the objective physiological measurements to avoid bias towards one data source.

The investigated driving scenarios reveal a significant influence of driving modes (e.g., sporty) on the trust in AD. It may also be concluded that trust evolves with participants experience of the system. That happens even with a relatively high starting trust, hence indicating ongoing learning effects for those who interact with an AD system. The insights support the need for early engagement with AD systems and human-centred development of AD.

Contradicting to other studies, we discover that pre-existing experience of driving automation might initially reduce trust in the system. However, that gain of trust then rises faster than for drivers without preexisting experience of the technology.

Despite the effect of the disparity between the simulated and real-world on the drivers, simulator studies seem to have

an advantage in early-stage development, where the direction of the results is more important than their intensity.

The data quality suggests that a small study group engaged in a customized study of a high internal validity can achieve high-quality results, which is promising for future GSR testing in DDAW and the next step of GSR. The future work is likely to be based on subjecting biometric data to machine learning methods to perform an objective assessment of user experience with subjective ratings.

REFERENCES

[1] Affairs, Assistant Secretary for Public: System usability scale (sus) (2013), https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

[2] Aiello, M.A., Agarwal, P.K., Hilfinger, S.H., Grigorian, C.H., Kantaros, D.W.: 2017 connected cars & autonomous vehicles survey. Foley & Lardner LLP (24/10/2017), https://www.foley.com/en/insights/publications/2017/10/2017-connected-cars--autonomous-vehicles-survey

[3] AVL List GmbH: Avl drive™ 4, https://www.avl.com/-/avl-drive-4-

[4] AVL List GmbH: Avl model.connect™, https://www.avl.com/de/-/model-connect-

[5] AVL List GmbH: Avl vsm™ vehicle simulation, https://www.avl.com/-/avl-vsm-4-

[6] Benedikt, M., Watzenig, D., Zehetner, J., Hofer, A.: Nepce - a nearly energy-preserving coupling element for weak-coupled problems and co-simulations. In: Idelsohn, S.R. (ed.) Coupled problems 2013. CIMNE, Barcelona (2013)

[7] Boucsein, W., Backs, W.: The psychophysiology of emotion, arousal, and personality: Methods and models. In: Duffy, V.G. (ed.) Handbook of digital human modeling, pp. 627–644. Human factors and ergonomics, CRC Press (2009). https://doi.org/10.1201/9781420063523-44

[8] Bracha, H.S.: Freeze, flight, fight, fright, faint: adaptationist perspectives on the acute stress response spectrum. CNS spectrums 9(9), 679–685 (2004). https://doi.org/10.1017/s1092852900001954

[9] Clement, P., Danzinger, H., Quinz, P., Hillbrand, B., Hartavi, A.E., Kasikci, K.Z.: Assessment concept for trustvehicles. In: Watzenig, D., Schicker, L.M. (eds.) Enhanced Trustworthiness and End User Acceptance of Conditionally Automated Vehicles in the Transition Period, pp. 107–128. Lecture Notes in Intelligent Transportation and Infrastructure, Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-60861-3_6

[10] Dimitrakopoulos, G., Uden, L., Varlamis, I.: The future of intelligent transport systems. Elsevier, Amsterdam (2020)

[11] EU: General data protection regulation eu3016/0679: Gdpr (04/05/2016), https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679

[12] EU: General safety regulation eu2019/2144: Gsr (16/12/2019), https://eur-lex.europa.eu/eli/reg/2019/2144/oj

[13] EURO NCAP: Euro ncap assessment protocol - sa - v9.0.3, https://cdn.euroncap.com/media/58229/euro-ncap-assessment-protocol-sa-v903.pdf

[14] EuroStat: International standard classification of education (isced), https://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED)

[15] FFG: Anti-coriolis-training in motion optimized simulators

[16] FFG: Integrating mental training techniques intro simulator training

[17] FFG: Proactive safety behaviour under stress and safety culture in air traffic management

[18] Gasser, T.M., Seeck, A., Smith, B.W.: Rahmenbedingungen für die fahrerassistenzentwicklung. In: Winner, H., Hakuli, S., Lotz, F., Singer, C. (eds.) Handbuch Fahrerassistenzsysteme, pp. 27–54. ATZ / MTZ-Fachbuch, Springer Vieweg, Wiesbaden (2015). https://doi.org/10.1007/978-3-658-05734-3_3

[19] Graham, F.K., Clifton, R.K.: Heart-rate change as a component of the orienting response. Psychological bulletin 65(5), 305–320 (1966). https://doi.org/10.1037/h0023258

[20] Grier, R.A., Bangor, A., Kortum, P., Peres, S.C.: The system usability scale. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57(1), 187–191 (2013). https://doi.org/10.1177/1541931213571042

[21] Helldin, T., Falkman, G., Riveiro, M., Davidsson, S.: Presenting system uncertainty in automotive uis for supporting trust calibration in autonomous driving. In: Terken, J. (ed.) Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Eindhoven, 2013, [October 28 - 30, 2013]. pp. 210–217. ACM, New York, NY (2013). https://doi.org/10.1145/2516540.2516554

[22] Imotions Publish: Pupil labs core - imotions (12/02/2021), https://imotions.com/hardware/pupil-labs-glasses/

[23] Index, The NASA TLX Tool: Task Load: Tlx @ nasa ames - home (19/04/2021), https://human-factors.arc.nasa.gov/groups/TLX/

[24] Jian, J.Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics 4(1), 53–71 (2000). https://doi.org/10.1207/S15327566IJCE0401_04

[25] Lee, J.D., Liu, S.Y., Domeyer, J., DinparastDjadid, A.: Assessing drivers' trust of automated vehicle driving styles with a two-part mixed model of intervention tendency and magnitude. Human factors 63(2), 197–209 (2021). https://doi.org/10.1177/0018720819880363

[26] Macher, G., Armengaud, E., Kreiner, C., Brenner, E., Schmittner, C., Ma, Z., Martin, H., Krammer, M.: Integration of security in the development lifecycle of dependable automotive cps. In: Management Association, I.R. (ed.) Research anthology on artificial intelligence applications in security, pp. 101–142. IGI Global, Hershey, Pennsylvania (701 E. Chocolate Avenue, Hershey, Pennsylvania, 17033, USA) (2021). https://doi.org/10.4018/978-1-7998-7705-9.ch006

[27] Macher, G., Druml, N., Veledar, O., Reckenzaun, J.: Safety and security aspects of fail-operational urban surround perception (fusion). In: Katsaros, P. (ed.) Model-Based Safety and Assessment, LNCS Sublibrary, vol. 11842, pp. 286–300. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-32872-6_19

[28] Molnar, L.J., Ryan, L.H., Pradhan, A.K., Eby, D.W., St. Louis, R.M., Zakrajsek, J.S.: Understanding trust and acceptance of automated vehicles: An exploratory simulator study of transfer of control between automated and manual driving. Transportation Research Part F: Traffic Psychology and Behaviour 58, 319–328 (2018). https://doi.org/10.1016/j.trf.2018.06.004

[29] Polar USA: Polar h10 — heart rate monitor chest strap — polar usa (19/04/2021), https://www.polar.com/us-en/products/accessories/h10_heart_rate_sensor

[30] SAE International: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, http://standards.sae.org/J2016201806

[31] Sahami, S., Sayed, T.: How drivers adapt to drive in driving simulator, and what is the impact of practice scenario on the research? Transportation Research Part F: Traffic Psychology and Behaviour 16, 41–52 (2013). https://doi.org/10.1016/j.trf.2012.08.003

[32] Schmittner, C., Macher, G.: Automotive cybersecurity standards - relation and overview. In: Romanovsky, A., Troubitsyna, E., Gashi, I., Schoitsch, E., Bitsch, F. (eds.) COMPUTER SAFETY, RELIABILITY, AND SECURITY, Lecture Notes in Computer Science, vol. 11699, pp. 153–165. SPRINGER (2019). https://doi.org/10.1007/978-3-030-26250-1_12

[33] Veledar, O.: New Business Models to Realise Benefits of the IoT Technology within the Automotive Industry. Ph.D. thesis, Unpublished (2019). https://doi.org/10.13140/RG.2.2.29671.39847

[34] VIRES: Vtd virtual test drive, https://www.mscsoftware.com/de/virtual-test-drive

[35] Wachenfeld, W., Winner, H.: The release of autonomous vehicles. In: Maurer, M., Gerdes, J.C., Lenz, B., Winner, H. (eds.) Autonomous Driving, pp. 425–449. Springer Berlin Heidelberg, Berlin, Heidelberg and s.l. (2016). https://doi.org/10.1007/978-3-662-48847-8_21

[36] Wikström, N., Parrilla, A.F., Jones, S.J., Grauers, A.: Energy-efficient cooperative adaptive cruise control with receding horizon of traffic, route topology, and traffic light information. SAE International Journal of Connected and Automated Vehicles 2(2) (2019). https://doi.org/10.4271/12-02-02-0006