In UX We Trust

Investigation of Aesthetics and Usability of Driver-Vehicle Interfaces and Their Impact on the Perception of Automated Driving

Anna-Katharina Frison[*†]
Technische Hochschule Ingolstadt
Ingolstadt, Germany
anna-katharina.frison@thi.de

Philipp Wintersberger[*†]
Technische Hochschule Ingolstadt
Ingolstadt, Germany
philipp.wintersberger@thi.de

Andreas Riener[†]
Technische Hochschule Ingolstadt
Ingolstadt, Germany
andreas.riener@thi.de

Clemens Schartmüller[†]
Technische Hochschule Ingolstadt
Ingolstadt, Germany
clemens.schartmueller@thi.de

Linda Ng Boyle
University of Washington
Seattle, WA, USA
linda@uw.edu

Erika Miller
Colorado State University
Fort Collins, CO, USA
erika.miller@colostate.edu

Klemens Weigl[‡]
Technische Hochschule Ingolstadt
Ingolstadt, Germany
klemens.weigl@thi.de

ABSTRACT

In the evolution of technical systems, freedom from error and early adoption plays a major role for market success and to maintain competitiveness. In the case of automated driving, we see that faulty systems are put into operation and users trust these systems, often without any restrictions. Trust and use are often associated with users' experience of the driver-vehicle interfaces and interior design. In this work, we present the results of our investigations on factors that influence the perception of automated driving. In a simulator study, N=48 participants had to drive a SAE level 2 vehicle with either perfect or faulty driving function. As a secondary activity, participants had to solve tasks on an infotainment system with varying aesthetics and usability (2x2). Results reveal that the interaction of conditions significantly influences trust and UX of the vehicle system. Our conclusion is that all aspects of vehicle design cumulate to system and trust perception.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Interactive systems and tools**;

KEYWORDS

automated driving systems; user experience; UX; trust; distrust; SAE J3016; aesthetic; reliability

ACM Reference Format:
Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Linda Ng Boyle, Erika Miller, and Klemens Weigl. 2019. In UX We Trust: Investigation of Aesthetics and Usability of Driver-Vehicle Interfaces and Their Impact on the Perception of Automated Driving. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300374

[*]Both first and second author contributed equally
[†]Also with Johannes Kepler University.
[‡]Also with Katholische Universität Eichstätt-Ingolstadt.

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300374

1 INTRODUCTION

The development of technology has focused on supporting individual mobility and satisfying human desire for autonomy, which is one of the most important psychological needs [65]. Each new invention – from the wheel over horse carriages, steam driven railways, up to the automobile – fundamentally changed our daily life and the societies we live in. Automated vehicles (AVs) are a major next step in this evolution. AVs promise several benefits, such as less congestion and pollution, higher safety, as well as more leisure time and enhanced mobility for diverse target groups (children, elderly, impaired) [51]. However, all these advantages cannot be delivered instantly. Fully automated "level 5" vehicles [12] that can operate in all circumstances are not expected on the market before 2030 or 2040 [50]. In the meantime, automated driving systems (ADSs) with lower levels of automation are being gradually introduced. During this era of mixed traffic (co-existence of vehicles at different levels of automation), drivers must be able to cope with automation limitations and act either as monitoring (level 2) or fallback (level 3) authority [13]. Monitoring over extended periods of time is a challenge, even for *"highly motivated human beings"* (c.f., irony of automation [1]). This is particularly relevant as drivers, who are not necessarily well trained domain experts, are expected to operate a safety critical system in potentially dangerous environments [76]. Recent incidents with AVs (such as the fatal accidents with Tesla Autopilot or the Uber self-driving taxi [73]) confirm that AV technology is highly susceptible to overreliance/overtrust [19]. Since this problem is already well known from research on driver assistance systems such as adaptive cruise control [16], trust calibration for AVs is an issue receiving a great deal of attention [31, 40, 58, 77]. Ideally, users' trust levels should appropriately "match an objective measure of trustworthiness", e.g., system performance (reliability/ability/predictability to achieve its goals) [45, 57]. However, reality is much more complex: trust in automation has many dimensions and is influenced by a variety of aspects, including aesthetics, design, and other factors of user experience (UX) [32]. Conversely, also UX is impacted by users' trust in a system. For vehicle manufacturers, this results in a nearly unsolvable paradox. On the one hand, they should design their systems in a way that can prevent overtrust/overreliance. On the other hand, they must maximize UX qualities of their vehicles to maintain competitiveness.

In the context of AVs, how the partly overlapping constructs of UX and trust actually influence each other is widely unknown. Vehicles consist of many subsystems and design aspects that may all contribute to users' overall assessment of both constructs and, further on, may cause so called "halo effects" [28, 49, 55, 67, 69, 71]. Hence, this paper aims at revealing how both UX and trust influence each other, and how proper design could simultaneously support UX qualities while preventing miscalibrated trust. To evaluate this, we conducted a driving simulator study, that, to the best of our knowledge, combines for the first time relevant parameters of both UX (usability, aesthetics) and trust (system performance/reliability) in a single experiment. In our study, users had to complete several tasks in an in-vehicle infotainment system (IVIS) while safely operating an AV in SAE level 2. We utilized a mixed-model design that varies the reliability of an ADS as a between-subjects factor. For the within-subjects factors, we varied pragmatic (representing usability) and hedonic (representing aesthetics) qualities of the IVIS. For a holistic evaluation of UX, trust, affect, and psychological need fulfillment, we applied a triangulation of subjective (AttrakDiff mini [29], PANAS short [74], Trust Scale [34], Need Scale [64], semi-structured interviews) and objective (galvanic skin response, braking behavior) measures. For all subjective measures, we emphasized participants to assess the AV as a whole and not distinguish between subsystems. The results of our study give insights in how the stream of experiences combining performance, usability, and aesthetics of different vehicle subsystems correlate and influence each other.

Figure 1: Study setup showing the driving scenario and an example of the IVIS used to investigate the interaction between UX and trust.

2 UX AND TRUST: THEORY AND PRACTICE

In the HCI community, UX and trust research are two areas which are often considered in isolation. However, similarities between both constructs cannot be denied and, thus, we propose to consider them in a holistic way.

Similarities

Trust in automation can be defined as *"attitude that an agent will achieve an individual's goals in a situation characterized by uncertainty and vulnerability"*, and is built upon analytic, analogical, and affective processes [45]. Trust is sensitive to individual traits (such as age, personality, etc.) and states (self-confidence, emotional state, etc.), properties of the automation (complexity, task difficulty, etc.), as well as design features (appearance, ease of use, communication style, etc.)

[32], and is the result of processes happening before (*"dispositional trust"*), during *("situational trust")*, and after *("learned trust")* system interaction [32]. In contrast, UX can (according to ISO 9241/210) be defined as a *"person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service"*. Thereby, experience can *"occur before, during and after use"*, and *"[...] is a consequence of brand image, presentation, functionality, system performance, interactive behaviour and assistive capabilities of the interactive system, the user's internal and physical state resulting from prior experiences, attitudes, skills and personality, and the context of use"*.

Although trust and UX have separate definitions, they seem to be influenced by similar factors and processes. Hence, it is not surprising that trust is a mentioned (however, not yet focused) construct in UX theory literature. The term trust is regarded as a component of UX [44], users' personal quality of experience [79] or as (context-dependent [35]) perceived value [66]. Desmet et al. [15] mention trust within their general set of 25 emotions relevant in human-product interaction. Although trust is not an emotion itself, a product can help users to feel confident and courageous if it is perceived as trustworthy. Thus, designers need to decide which psychological needs they want to fulfill. Distler et al. [17] revealed the need of security as one of the most important needs for AVs (in the driving domain, the term "need for safety" would presumably fit better than "security", but for reasons of consistency we stick to the original formulation provided by [64] throughout the paper). In order to fulfill this need, a specific form of interaction has to be selected which aims at expressing trustworthiness and thereby triggers trust [26, 37, 41, 46]. In this sense, trust can be regarded as subjective sentiment and evaluative feeling dependent on the fulfillment of users' higher goals, such as the psychological need of security [23, 27]. To provide examples, Väätäjä et al. [70] include trust as item in the AttrakWork questionnaire to measure a products' hedonic quality and Roedel et al. [62] chose trust as relevant UX factor when evaluating user acceptance and experience of ADSs at different levels of automation. Hence, a question that could arise in this regard, especially considering that both constructs are discussed as very broad, fuzzy and hard to understand [32, 43, 60, 78], is, whether or not trust and UX can be considered the same in a specific context?

Differences

The main difference becomes visible when looking at the goals both constructs aim to achieve. UX research tries to maximize the quality of interaction by satisfying psychological needs and thereby providing pragmatic and hedonic quality [23, 25]. For designers, there is no upper limit – the more these qualities are supported, the better. Thus, previous research focused on the impact of visual aesthetics, usability, and branding on users' perceived trustworthiness, predominantly in the area of e-commerce systems [18, 37, 47] and websites [49]. These studies aimed to increase users' perceived trustworthiness and, consequently, enhance UX. In trust research however, maximizing trust is not the major goal. Here, the challenge is to precisely adjust users' subjective trust levels to a systems' actual performance ("calibration of trust" [57]) while taking the operational and environmental context into account [45]. Thus, although trust may need to be raised in many situations, an upper limit should not be exceeded to prevent users from overreliance. In the domain of automated driving, recent studies addressing trust can broadly be divided into two areas. Those dealing with distrust to reduce automation disuse, and those that address the problem of overtrust/overreliance to prevent misuse [57]. For both issues, various resolution strategies have been proposed. Trust may be raised by increasing system transparency, using various techniques such as why-and-how information [38], symbolic representation [31], augmented reality [77] or anthropomorphic agents [39, 75]. An often proposed solution to deal with overtrust is the provision of uncertainty displays in different forms and modalities [3, 40, 58]. In this context, a problem that we see in many trust studies is that a distinction between the two constructs (trust and UX) is not made. For example, was the aim of an experiment actually to address trust/reliance or were mainly UX aspects evaluated which potentially overlap with trust?

Research Opportunities

Lindgaard et al. [49] claim that so called "halo effects" are a reason for the interrelation of usability, aesthetics, and trust in websites. These effects emerge from the paradox of "what is beautiful is usable" [67] or "I like it, it must be good on all attributes" [71], already mentioned by [24, 28, 59]. The interference model [28, 71] proved the existence of evaluative consistency (i.e., "halo effects"), which assumes that users interfere unavailable attributes from a general value to keep their overall judgment consistent. Hence, there is an indirect link between beauty which leads to goodness and, with it, pragmatic quality. In contrast, a probabilistic consistency is a conceptually or causally linked judgment (high aesthetics expects a high perceived hedonic quality). According to this, Tuch et al. [69] identified negative affects, such as frustration from poor usability, as a mediator variable that potentially decreases perceived aesthetics. Further, Minge et al. [55] differentiate between pragmatic "halo effects", where usability impacts perceived visual attractiveness, and hedonic "halo effects", where visual aesthetics influences perceived usability. Consequently, we wonder if trust in automation can be investigated in the absence of UX to draw useful conclusions. A central question that arises is, how the two constructs are correlated in the context of AVs? Similiar to [49], we expect halo effects of aesthetics and usability as biasing factors for trust, what could become highly relevant for the future implementation of automated driving technology.

3 USER STUDY

We conducted a driving simulator study to investigate the interaction (potential correlation and "halo effects" of UX and trust) between an ADS's performance/reliability and relevant UX factors (usability/aesthetics) of in-vehicle interfaces, as well as their effect on the perception of AVs in general; aiming to answer the following research questions:
RQ1: How does IVIS design (usability and aesthetics) affect UX of AVs with varying system performance?
RQ2: How does IVIS design (usability and aesthetics) affect users' trust in AVs with varying system performance?
RQ3: Is there a correlation between UX and trust in AVs?

Experimental Design

We applied a full factorial mixed-model design varying the performance of the ADS as between-subjects factor, and aesthetics and usability of the IVIS as within-subjects factor (each on two levels). Each participant had to perform various tasks on four different IVISs that represented all combinations of usability (good/bad) and aesthetics (nice/ugly).

Study Setup

The experiment was conducted in a high-fidelity driving simulator (remodeled VW Golf on hexapod platform) and an IVIS on a tablet PC installed on top of the center console (see Figure 1).

Driving Scenario. We simulated an AV at SAE level 2 (i.e., combination of longitudinal and lateral control) driving on a 2-lane highway using IPG CarMaker, inspired by the setting used in [3]. The AV drove with a constant speed of 120km/h on the left lane and was confronted with 12 lead vehicles driving at lower speed (70km/h). In such a situation, the ADS detected the lead vehicle and reduced the speed to prevent a crash (similar to an ACC system). As soon as the ego vehicle slowed down to 70km/h, the lead vehicle performed a lane change to the right, allowing the ego vehicle to accelerate again to the target speed. In the high-performance condition (group A), all 12 lead vehicles were successfully detected (thus, no manual interventions were necessary). In the low-performance condition (group B), the ADS (randomly) failed to detect the lead vehicle in 3 out of the 12 cases (75% reliability), generating the need for interventions – participants thus had to brake manually to prevent a crash (however, they never had to manually engage in lateral control).

In-Vehicle Infotainment System. We implemented four variants of IVISs in HTML/Javascript on a 10.2" tablet (Google Pixel C). The IVISs consisted of a main navigation and three typically available subsystems (a phone/call screen including a list of contacts, a media player including a collection of albums/songs as well as different radio stations, and a climate control), see Figure 3.

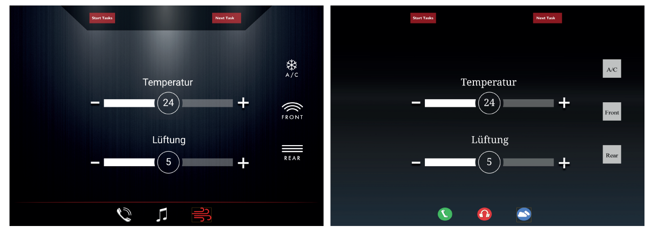

Figure 3: A/C menu of the nice (left) and ugly (right) IVIS.

The visual design was selected from a set of examples created by groups of undergraduate students during a design class. Students were provided a specific menu/navigation structure and instructed to create an IVIS skin. All designs were evaluated using the UEQ [42] on a 7-point semantic differential scale from -3 (negative) to +3 (positive) with at least 5 participants. We utilized the results of the subscale "Attractiveness (Att-UEQ)" and selected the IVISs with the best and worst values. While the nice design has a mean value of ATT-UEQ=1.92 (excellent with respect to the UEQ benchmark dataset [63]), the ugly design shows mean value of only ATT-UEQ=0.45 (bad compared to the benchmark). This process aimed as guidance to confirm our subjective selection of a nice and ugly IVIS, however, was no controlled experiment. To provide a potentially "bad" usability, we followed the definition provided in ISO 9241-11 that states usability to be the *"extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* [33]. Thus, we chose to manipulate the IVISs reliability by semi-randomly calculating the chance for a successful button-press action, where at least two and at most 8 clicks were required for a successful action.

Participants and Procedure

In total, 48 participants (16 female, 32 male) aged between 19 and 26 (M_{age}= 22.09, SD_{age} = 1.89) years, all undergraduate students, voluntarily participated in the experiment. Each participant was assigned to either group A (high ADS performance) or group B (low ADS performance), potential differences between the groups considering gender and age were counterbalanced. No participant had to be excluded due to simulator sickness or technical problems. After completing a short questionnaire assessing demographics, each subject conducted a 3-minute test drive to become familiar with the AV. Then, we instructed participants that they

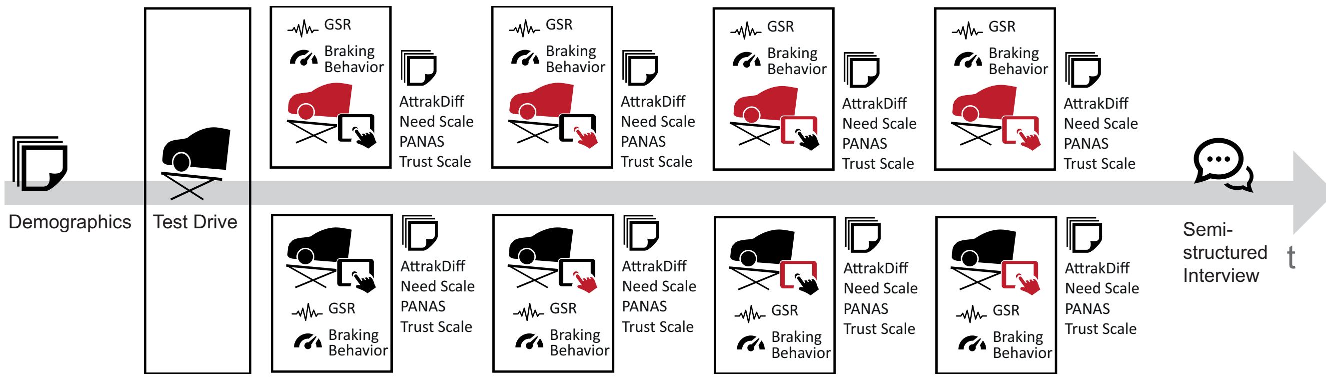

Figure 2: Study procedure: The top row represents the drives with low, the bottom row drives with high ADS performance. The red color indicates decreased qualities (finger: usability, tablet: aesthetics, driving simulator: performance).

will experience four different types of AVs with different IVISs. We further told them that manual braking interventions could be necessary due to automation failures, and that safely completing the drive has the highest priority. Afterwards, participants experienced four consecutive 5-minute lasting trips while experiencing the 4 different IVISs (in randomized order). Within each condition, participants had to complete seven tasks on the IVISs with two levels of complexity. Easy tasks consisted of a single instruction only (such as "call John"), while complex tasks required participants to remember multiple steps (such as "switch to Radio Disney Channel and adjust the volume to 8"). The task instructions were presented auditory (pre-recorded sound files). Successful completion of a task was indicated with a notification sound and the next task was issued 35 seconds afterwards. In case all seven tasks were completed before finishing the 5-minute lasting drive, the experimental condition was stopped earlier. The selection of tasks from the set was randomized over the conditions, and quasi-randomized within the scenarios (each task was only presented once during the entire experiment). After each condition, participants had to complete a survey including a set of different standardized scales to assess trust and UX in the AV (see Figure 2), whereby we repeatedly instructed them to assess the AV as a whole, single system based on their experiences. Additionally, a short semi-structured interview with all participants was conducted after the experiment to reveal further insights into their thoughts and attitudes. The whole experiment lasted approx. 90 minutes for each participant.

Data Collection

To be able to evaluate the proposed research questions, we triangulate a set of subjective and objective measures derived from established theory as emphasized in the following.

Subjective Measures. To assess UX and trust, we utilized multiple subjective scales. We used the AttrakDiff mini [29] with a 7-point semantic differential scale ranging from 0 (low) to 6 (high). Thereby, the subscale attractiveness (ATT), consisting of two items for beauty and goodness, assesses the overall perception combining both pragmatic (PQ) and hedonic quality (HQ). Since for all subscales Cronbachs' α resulted in acceptable values (> .60, see Table 1), we calculated mean scale values. All UX qualities are intercorrelated (Pearson's correlation coefficient), ranging from r=.412 to r=.880 across all conditions. HQ and PQ showed least (r<.60), HQ and ATT highest intercorrelations (r>.60). As UX is also dependent on the satisfaction of psychological needs [27, 30], we further utilized the need scale (same version as used in [27] with 7-point Likert scale) and focused on the needs most relevant in the context of AVs: autonomy (AUT), competence (COM), stimulation (STI), and security (SEC) [21]. Also here, reliability of all subscales was acceptable ($\alpha > .70$, see Table 1). Intercorrelation between the subscales across all conditions ranged from r=.26 to r=.81. Further, system interaction leads to particular (positive and negative) emotions [14] resulting from need fulfillment [27, 30]. Thus, we included the short version of PANAS also with a 7-point Likert scale [53, 74]. PA and NA did not correlate (r <.12) and reliability of all subscales was acceptable ($\alpha > .70$, see Table 1).
To evaluate subjective trust we used the trust scale provided by Jian et al. [34]. This scale consists of two subscales for trust (T) and distrust (DT) (7-point Likert) and is widely used to assess trust in automation or robotic systems [36, 58]. Also here, Cronbachs' α resulted in acceptable values while T and DT showed a negative correlation ($r > -.80$).

Objective Measures. Galvanic Skin Response (GSR) is commonly used as an indicator for the sympathetic nervous system. Changes in skin conductance have been linked to arousal [10, 11], (cognitive) workload [6, 9], usability [48], user experience [22], but also trust [56]. Signal peaks, so called Skin Conductance Responses (SCRs), indicate such activation while the general signal level is subject to bias

Dep. Variable	Items	Cronbach's α	Ref.
UX Qualities			
Attractiveness (ATT)	2 (Beauty and Goodness)	.65	[29]
Pragmatic Q. (PQ)	4	.77	[29]
Hedonic Q. (HQ)	4	.79	[29]
Needs			
Autonomy (AUT)	3	.84	[27, 64]
Competence (COM)	3	.86	[27, 64]
Stimulation (STI)	3	.83	[27, 64]
Security (SEC)	2	.77	[27, 64]
Affect			
Positive (PA)	5	.75	[74]
Negative (NA)	5	.85	[74]
Trust			
Trust (T)	6	.91	[34]
Distrust (DT)	5	.87	[34]

Table 1: Summary of subjective methods employed.

by individual differences, room temperature, etc. [6]. We utilized a professional 500 Hz physiological measurement system from g.tec medical engineering (www.gtec.at) and attached two skin electrodes to the volar (inner) middle phalanges (muscle limbs) of the non-dominant hand's middle and ring fingers (see guidelines by [7, 8]). Since GSR is sensitive to motion artifacts, we instructed participants to behave naturally but also to prevent waving their hand excessively. We used Ledalab for Matlab [4] to extract all SCRs since the implemented Continuous Decomposition Analysis (CDA) is supposed to be more robust at discriminating single SCRs than traditional peak-detection methods [5]. For the evaluation, we utilized the number of SCRs, which is argued to be less affected by individual differences and other forms of bias [6]. To evaluate driving behavior, we recorded participants' brake pedal actuation and calculated three parameters – the number of brakes representing the quantity of manual interventions, the average duration of a brake pedal actuation, and the average brake intensity (on a scale from 0 to 1).

4 RESULTS

In the following we present a detailed analysis of the collected data with respect to our research questions (all results with $p < .05$ are reported as statistically significant). Since tests for normality (Shapiro-Wilk's, $p > .05$), marginal existence of outliers, and homogeneity of error variances assessed by Levene's test ($p > .05$) were passed for all dependent variables (except for driving performance, see Table 1), parametric tests were applied. We performed three-way mixed ANOVAs with the independent variables ADS performance as between-subjects, and IVIS usability and aesthetics as within-subjects factors. As the collected driving performance measures did not follow a normal distribution, non-parametric tests (Mann-Whitney-U tests for the between-, and Wilcoxon Signed-Rank tests for the within-subject factors) were applied. To analyze correlations between the subjective constructs of UX and trust, we conducted Pearson's bivariate correlation analyses.

User Experience (RQ1)

To answer RQ1, we analyzed the data of UX (UX Qualities, Needs and Affect) scales as well as the objective data on participants' arousal given by GSR. Concerning multivariate tests statistics, we utilized Pillai's Trace.

UX Qualities. Multivariate tests evaluating the impact of ADS performance, IVIS aesthetics and usability on participants' perception of product quality (measured by AttrakDiff) reveals no significant main effect for the between-subject factor ADS performance ($V = .21, F(5, 42) = 2.24, p = .068$). However, separate univariate ANOVAs on the outcome variables show a significant effect for pragmatic quality (PQ, $F(1, 46) = 8.62, p = .005, \eta^2 = .16$). Results for high ADS performance were perceived as significantly better than for low performance conditions. Ratings for attractiveness (ATT, Goodness and Beauty) and hedonic quality (HQ) did not differ significantly (see Table 2). Additionally, multivariate tests reveal that the overall perceived system quality significantly differs regarding IVIS usability ($V = .44, F(5, 42) = 6.47, p < .001$). Univariate tests confirm a significant effect for ATT ($F(1, 46) = 14.67, p = .001, \eta^2 = .24$). Regarding the items "Goodness" and "Beauty" separately, there is only a significant effect on "Goodness" ($F(1, 46) = 25.45, p < .001, \eta^2 = .36$). Also PQ ($F(1, 46) = 25.36, p < .001, \eta^2 = .36$) and HQ ($F(1, 46) = 10.60, p = .002, \eta^2 = .19$) differed significantly. Thus, systems with good IVIS usability were perceived better than those with bad IVIS usability across all conditions. Further, we can report a significant main effect for the within-subject factor IVIS aesthetics ($V = .57, F(5, 42) = 11.24, p < .001$). Univariate tests reveal significant effects for ATT ($F(1, 46) = 50.22, p < .001, \eta^2 = .52$). Here, both items, "Goodness" ($F(1, 45) = 20.22, p < .001, \eta^2 = .30$) and "Beauty"($F(1, 46) = 58.23, p < .001, \eta^2 = .56$), show significant effects. Also PQ ($F(1, 46) = 28.29, p < .001, \eta^2 = .38$) and HQ ($F(1, 46) = 52.44, p < .001, \eta^2 = .53$). Thus, across all conditions the nice IVIS was rated better than the ugly IVIS. Moreover, our data confirms the inference model [28, 71] – better aesthetics leads to a significantly higher ratings for goodness and therewith higher ratings for PQ, and not only beauty (evaluative consistency). However, no two or three-way interaction effects could be revealed.

Needs. For users' need fulfillment of Autonomy (AUT), Competence (COM), Stimulation (STI), and Security (SEC), we can report a significant effect for ADS performance regarding the multivariate test statistic ($V = .22, F(4, 43) = 8.09, p = .025$). Univariate tests reveal only significant differences in participants' need of SEC ($F(1, 46) = 12.88, p = .001, \eta^2 = .22$), which was less fulfilled in the group with the low ADS performance. Multivariate tests show a significant main effect for IVIS usability ($V = 0.24, F(4, 43) = 3.33, p = .018$), univariate tests resulted in a significant decrease of SEC in case

				95% Confidence Interval	
Dep. Variable	**Ind. Variable**	M	SD	lower	upper
	ADS performance				
PQ	high	4.17	0.18	3.81	4.53
	low	3.42	0.18	3.06	3.79
	IVIS usability				
ATT	good	3.36	1.01	3.07	3.65
	bad	2.96	1.14	2.62	3.30
↪ Goodness	good	3.74	1.23	3.39	4.09
	bad	2.96	1.45	2.54	3.37
PQ	good	4.10	0.92	3.86	4.34
	bad	3.49	1.14	3.17	3.81
HQ	good	2.97	0.83	2.72	3.21
	bad	2.68	0.89	2.42	2.94
	IVIS aesthetics				
ATT	nice	3.72	1.06	3.40	4.04
	ugly	2.60	1.10	2.24	2.97
↪ Beauty	nice	3.72	1.34	3.39	4.05
	ugly	2.24	1.37	3.39	4.09
↪ Goodness	nice	3.72	1.34	3.34	4.10
	ugly	2.98	1.38	2.58	3.38
PQ	nice	4.05	0.97	3.78	4.32
	ugly	3.54	1.04	3.26	3.82
HQ	nice	3.38	0.80	3.14	3.61
	ugly	2.27	1.10	1.95	2.59

Table 2: Significant UX Quality Values.

of bad IVIS usability ($F(1, 46) = 7.43, p = .009, \eta^2 = .14$) and COM ($F(1, 46) = 9.54, p = .003, \eta^2 = .17$). Further, also for the within-subject factor IVIS aesthetics, a significant main effect could be revealed ($V = .24, F(4, 43) = 3.38, p = .017$). Regarding univariate tests we can observe effects for the need of STI ($F(1, 46) = 12.12, p = .001, \eta^2 = .21$), AUT ($F(1, 46) = 5.22, p = .027, \eta^2 = .10$), SEC ($F(1, 46) = 6.26, p = .016, \eta^2 = .12$) and COM ($F(1, 46) = 6.08, p = .017, \eta^2 = .12$). Thereby, all these needs are significantly less fulfilled when driving in an AV with ugly IVIS (see Table 3 for means). Here, data analysis did not reveal any two- or three-way interaction effects.

				95% Confidence Interval	
Dep. Variable	**Ind. Variable**	M	SD	lower	upper
	ADS performance				
SEC	high	3.24	1.20	2.79	3.69
	low	2.10	0.25	1.65	2.56
	IVIS usability				
SEC	good	2.82	1.34	2.47	3.18
	bad	2.52	1.23	2.20	2.84
COM	good	3.13	1.37	2.73	3.52
	bad	2.79	1.36	2.40	3.18
	IVIS aesthetics				
AUT	nice	2.48	1.34	2.09	2.86
	ugly	2.30	1.41	1.89	2.71
STI	nice	2.77	1.12	2.45	3.09
	ugly	2.40	1.27	2.04	2.76
SEC	nice	2.82	1.28	2.48	3.16
	ugly	2.53	1.30	2.19	2.87
COM	nice	3.07	1.26	2.71	3.44
	ugly	2.84	1.43	2.43	3.26

Table 3: Significant Need Values.

Affect. Participants' positive (PA) and negative affect (NA) revealed a significant main effect for ADS performance, ($V = .36, F(2, 45) = 11.33, p < .001$). A look at univariate tests revealed that NA for the low ADS performance is significantly higher than for the high ADS performance condition($F(1, 46) = 23.14, p < .001, \eta^2 = .34$), while PA was not affected. Regarding the within-subject factor IVIS usability, we can observe similar results. Multivariate tests reveal a significant main effect ($V = .16, F(2, 45) = 4.38, p = .018$), however, also here only NA showed differences in case IVIS usability is bad ($F(1, 46) = 7.26, p = .010, \eta^2 = .14$). Contrarily, IVIS aesthetics, which also has a significant main effect ($V = .21, F(2, 45) = 6.007, p = .005$), shows significant differences for both PA ($F(1, 46) = 4.24, p = .045, \eta^2 = .08$) and NA ($F(1, 46) = 8.63, p = .005, \eta^2 = .16$). Thereby, PA is slightly (but still significantly) higher for the nice IVIS aesthetics in contrast to the ugly IVIS variants. Again, no two or three-way interaction effects could be revealed (see Table 4 for means).

				95% Confidence Interval	
Dep. Variable	**Ind. Variable**	M	SD	lower	upper
	ADS performance				
NA	high	1.04	0.16	0.64	1.44
	low	2.40	0.23	2.00	2.80
	IVIS usability				
NA	good	1.58	1.26	1.27	1.89
	bad	1.86	1.22	1.57	2.15
	IVIS aesthetics				
PA	nice	2.98	0.98	2.69	3.27
	ugly	2.80	1.05	2.49	3.11
NA	nice	1.57	1.21	1.27	1.86
	ugly	1.87	1.27	1.56	2.18

Table 4: Significant Affect Values.

Arousal. Analysis of GSR data revealed a significant main effect for the within-subject factor IVIS usability ($F(1, 38) = 9.85, p = .003, \eta^2 = .21$). Bad IVIS usability leads to significantly more peaks, thus arousal, than good usability. We further can observe a two-way interaction effect for IVIS usability and ADS performance($F(1, 38) = 4.98, p = .032, \eta^2 = .12$). Descriptive statistic show that if ADS performance is low and IVIS usability is bad, participants are significantly more aroused than if ADS performance is high and IVIS usability is good. However, when ADS performance is low although the IVIS usability is good, the number of GSR peaks is also increasing. No further main effects for ADS performance or IVIS aesthetics, and also no further two- and three-way interaction effects could be revealed by our statistical analysis (see Table 5 for descriptive statistics).

				95% Confidence Interval	
Dep. Variable	**Ind. Variable**	M	SD	lower	upper
	IVIS usability				
Peaks	good	203.06	65.66	183.99	222.14
	bad	220.45	78.46	196.84	244.07
	ADS performance x IVIS usability				
Peaks	high & good	194.20	67.20	167.22	221.18
	high & bad	223.95	81.00	190.55	257.35
	low & good	211.93	63.00	184.95	238.90
	low & bad	216.95	77.71	183.55	250.35

Table 5: Significant Arousal Values.

Trust (RQ2)

To answer RQ2, we analyzed the subjective trust ratings as well as participants' braking behavior.

Trust Scale. Multivariate data analysis (using Pillai's Trace) of users' trust (T) and distrust (DT) revealed a significant main effect for ADS performance ($V = .29, F(2, 45) = 9.02, p = .001$). Univariate tests on the dependent variables show significant effects for T ($F(1, 46) = 18.07, p < .001, \eta^2 = .28$) and DT ($F(1, 46) = 15.09, p < .001, \eta^2 = .25$). While T is decreasing in conditions of low ADS performance, DT is increasing. Contrarily, T is increasing for high ADS performance and DT decreasing. We can report another main effect for IVIS usability ($V = .24, F(2, 45) = 7.12, p = .002$). Also here, significant effects for T ($F(1, 46) = 14.54, p < .001, \eta^2 = .24$) and DT ($F(1, 46) = 9.48, p = .003, \eta^2 = .17$) are visible. Descriptive data shows similar effects like for the between-subject factor ADS performance. Further, also IVIS aesthetics shows a significant main effect ($V = .22, F(2, 45) = 6.17, p = .004$). However, here only DT could be significantly decreased by a nice IVIS interface, ($F(1, 46) = 12.58, p = .001, \eta^2 = .22$); see Table 6).

				95% Confidence Interval	
Dep. Variable	**Ind. Variable**	M	SD	lower	upper
	ADS performance				
T	high	3.91	0.21	3.34	4.36
	low	2.55	0.24	2.09	3.00
DT	high	2.34	0.20	1.90	2.79
	low	3.56	0.24	3.11	4.01
	IVIS usability				
T	good	3.40	1.31	3.08	3.72
	bad	3.06	1.35	2.71	3.40
DT	good	2.80	1.26	2.48	3.12
	bad	3.10	1.31	2.76	3.44
	IVIS aesthetics				
DT	nice	2.81	1.26	2.49	3.13
	ugly	3.09	1.28	2.76	3.43

Table 6: Significant Trust Values.

Braking Behavior. Since braking data was not normal distributed we performed non-parametric tests. Mann-Whitney U tests with Bonferroni correction (α=.0125) were conducted to confirm expected differences in braking behavior between low and high ADS performance. All braking parameters are, across all IVIS conditions, significantly higher in conditions with low than with high ADS performance (see Table 7).

To compare the impact of the IVIS on braking behavior, we calculated separate Friedman tests for low and high ADS performance with Bonferroni correction (α=.008). The number of brake actions differs only significantly for the group of the low ADS performance ($\chi^2(3) = 11.04, p = .012$). Post-hoc analysis revealed significant differences only between good & nice and bad & nice ($p = .022$), which led to more brake actions. Further, also braking duration is significantly different in conditions with low ADS performance ($\chi^2(3) = 13.40, p = .004$).

		ADS performance		Test Statistic
	IVIS	Mdn (high)	Mdn (low)	Mann-Whitney-U test
Number	bad & nice	0	**5***	U = 510, z = 4.73, p < .001
	bad & ugly	0	5	U = 515, z = 4.83, p < .001
	good & nice	0	**7***	U = 530, z = 5.12, p < .001
	good & ugly	0	6	U = 499, z = 4.47, p < .001
Duration	bad & nice	0	**2.65***	U = 511, z = 4.72, p < .001
	bad & ugly	0	2.47	U = 491, z = 4.30, p < .001
	good & nice	0	**1.99***	U = 438, z = 3.16, p < .002
	good & ugly	0	**2.42***	U = 511, z = 4.70, p < .001
Intensity	bad & nice	0	.72	U = 533, z = 5.19, p < .001
	bad & ugly	0	.72	U = 536, z = 5.25, p < .001
	good & nice	0	.55	U = 515, z = 4.79, p < .001
	good & ugly	0	.65	U = 519, z = 5.19, p < .001

Table 7: Braking Behavior. Significances between variables are indicated by *.

Post-hoc analysis revealed a significant difference only between good & nice, which shows lowest braking duration median and bad & nice with the highest braking duration median (p = .005), and additionally between good & nice and good & ugly (p = .022). For braking intensity, no significant effects could be revealed.

User Experience x Trust (RQ3)

To evaluate a potential correlation between the constructs UX and trust we ran bivariate Pearson correlation analyses of averaged correlation-coefficients after Fisher's Z-Transformation (see Table 8). Thereby, we applied Bonferroni Correction and adjust the significance level to α =.016.

	Trust (T)	**Distrust (DT)**
UX Qualities		
ATT	.5*	-.48*
↪ Beauty	.27	-.25
↪ Goodness	.59*	-.57*
HQ	.37*	-.36*
PQ	.68*	-.66*
Needs		
AUT	.25	-.16
COM	.30*	-.18
STI	.29	-.25
SEC	.75*	-.74*
Affect		
PA	.06	.04
NA	-.79*	.8*

Table 8: Averaged correlations between measures after z-transformation. Significances are indicated by * (Bonferroni-corrected).

Correlations. Participants' product quality perceptions show correlations with the constructs trust (T) and distrust (DT, s. Table 8). Although the overall perceived attractiveness (ATT) and almost all sub components correlate positive with T and negative with DT, the sole perception of beauty does not correlate significantly with T or DT. Regarding correlations of participants' psychological needs, we can observe a significant positive correlation of the need for security (SEC) and T, and a negative correlation with DT. The need of competence

(COM) correlates positive with T. Moreover, only negative affect (NA) correlates negative with T and positive with DT. Arousal and the construct trust do not correlate across all conditions. Also, no correlation could be identified between arousal and braking behavior.

Semi-structured Interviews. Semi-structured interviews (translated from German) confirm a correlation between perceived pragmatic quality and trust. Thereby, also participants in group with low ADS performance expressed to trust the system with good IVIS usability most: *"I would trust most in the ADS with a running infotainment system. If this is running I can also concentrate on other things around because I know this works"* (P3, low ADS performance). Several participants mentioned the distraction from monitoring the ADS as reason for decreased comfort and trust in the condition with bad IVIS usability. For some participants the influence of usability and aesthetics on trust was conscious, e.g., *"the whole vehicle has to look appealing and of high-quality that I agree to drive automated. The whole concept needs to be harmonious."* (P13, high ADS performance) Others, in contrast could not identify why they trusted most in the ADS with the good and nice IVIS. For example, one participant in the low ADS performance condition rated the ADS with good and nice IVIS as most trustworthy, however, reasoned *"because the AV performed best here"*(P5, low ADS performance) – actually, automation performed equally good for a in all conditions he experienced. Participants experiencing high ADS performance expressed that their trust increased gradually from beginning of the experiment to the end: *"At the beginning I was nervous while solving the tasks and I looked always on the street. In the end I relied on that the ADS is working"* (P1, high ADS performance). Another participant stated: *"The longer I tested the system, the more I trusted in it. The system I trusted most was the AV used in the second drive (nice and good), the interface of the IVIS was the most beautiful. My overall experience was impacted by it, thus, I also trusted more in this AV"* (P21, high ADS performance).

5 DISCUSSION

In the following we discuss the RQs and derive implications for AV research and development. Regarding **RQ1**, all independent variables show influence on multiple UX qualities. Especially the large influence of visual design on UX regarding users' higher goals confirms results from previous studies investigating the "halo effect" of usability and aesthetics [28, 49, 55, 67, 69, 71] in the context of AD. As ADS performance solely affected pragmatic aspects and thereby only the negative affect (probabilistic consistency), we can assume objective system performance to be a hygiene factor [68] for UX. Experience is only negatively affected if high system performance cannot be achieved.

Regarding trust (**RQ2**), ADS performance led to different results for both trust and distrust, what is also visible for the within-subject factor usability (probabilistic consistency). Aesthetics affected only distrust (with respect to our study sample). Thus, trust cannot be increased by a good design only, however, distrust can be decreased. This can be regarded as evaluative consistency, as there is no direct relation. The mutual influence of the independent variables on subjective trust indicates that users hardly differentiate between (for the driving task) more (ADS performance) and less (IVIS) important subfunctions (what Lee and See refer to as "low functional specificity" [45]). Still, we see a clear connection of perception and actual behavior. When looking at driving behavior, we can see that when UX aspects were degraded, participants actuated the brakes longer, thus de-accelerated to lower speeds and drove more carefully (this statement can be made as braking intensity did not differ, thus longer braking actions with similar intensity consequently lead to lower driving speed).
Correlation analysis further confirmed the familiarity of both constructs (**RQ3**). All UX quality dimensions (beside the perception of beauty), the psychological need for security and negative affect were correlated with trust/distrust. The influence of usability/aesthetics on trust was further emphasized in semi-structured interviews, even though some participants were not conscious of the impact. Our results do also not rely on subjective data only. Obtained GSR data shows that impairment of ADS performance and usability led to significantly higher arousal. Considering our results, we suggest the following recommendations for researchers and designers of automated driving systems:

Creating Public Awareness about System Complexity. The mutual influence of all variables reveals a huge problem – halo effects and low functional specificity considering trust confirm that it is hard for users to (at least initially) assess an AV based on objective characteristics. This is a known issue for interactive products, however, for AVs, the resulting negative effects might be dramatic. For example, falsely inferring trustworthiness from design aspects due to evaluative consistency could quickly lead to hazardous situations, and the safety critical environment simply does not allow longer system exposure and real-life experiences mediating this effect later on. Public authorities and/or vehicle manufacturers must thus create awareness, for example by adopting teaching practices in driving schools, public campaigns, etc.

More Sophisticated Study Design and Evaluation. We highly recommend trust researchers to design studies addressing trust more carefully, especially regarding evaluation methods. We recommend including UX measurements into studies that rely on subjective trust scales to better distinguish the outcomes of the objective properties of trust (performance aspects) from design aspects. When evaluating HMI for trust calibration, we recommend using a minimalist design to reduce the influence of design aspects or, even better, evaluate the same concept with varying degree of aesthetics to see if the desired effects are independent of the actual implementation. UX researchers, on the other hand, should more carefully consider the consequences reporting trust in their studies, particularly when evaluating safety critical systems such as AVs. Instead of seeing trust "just as another factor of UX", we urge them to regard the concept of trust/reliance in relation to system capabilities, as well as the danger of overtrust.

ADS Performance as Hygienic Factor. UX and trust are impaired in all conditions of unreliable ADS performance. Thus, primary objectives should be to improve automation, and such improvements should become integral part of the user interface. As the need for security seems to be most relevant for ADSs [17], the success of AVs will be dependent on the introduction of higher levels of automation where monitoring is no more needed. Recent studies conducted at real test tracks indicate that many drivers are not capable of intervening in upcoming crash situations despite eyes on the road and hands on the wheel [72]. A valid strategy could be to not offer vehicles operating at SAE level 2, which is of interest to the automotive companies, but unfortunately, difficult to achieve given the imperfections of the existing technology.

Don't Sell a Wolf in Sheep's Clothing. Vehicle designers should carefully consider halo-effects and it must be prevented to give users the impression that systems perform better than they actually do. Theoretically, our results could suggest that systems should be designed with bad usability and low aesthetics to reduce the chance of overtrust. However, it is clear that vehicle manufacturers aim for maximizing UX qualities to maintain competitiveness and enthuse customers for their products. This is also necessary to achieve broad acceptance/proliferation of ADSs on the market. Thus, they urgently need to take other methods into account to better communicate performance aspects to users. Manufacturers of ADSs should immediately include solutions that have already been suggested to approach the problem – such as making their systems transparent for the users by communicating system decisions [38, 77] and uncertainties [3, 40] or behavioral measures to avoid misuse (such as preventing automation from being enabled in environments it was not designed for).

6 LIMITATIONS AND FUTURE WORK

The presented work has some limitations. As differences between age groups concerning ADS experience exist, which are in particular related to the need of security and trust [20], future research needs to address this issue by involving a more heterogeneous user group. Thereby, age, cultural background, or personality must be included to achieve more generalizable results (as suggested by [2]). Another limitation of our study is the simulation environment. Although many studies addressing trust are conducted with driving simulators [40, 56, 77], their results must be interpreted cautiously. Also the IVIS implemented on a tablet computer was only an example, and since we could reveal strong influence of non-performance based aspects (such as aesthetics), other interfaces present in our simulator might have influenced results too. Future work thus needs to build up on our results and conduct studies in real AV prototypes and in authentic road conditions. Further, the impact of in-vehicle technology that supports non-driving related tasks [61] but also of unobtrusive interfaces for trust calibration, like light designs [52, 54], should be looked at in detail.

7 CONCLUSION

In this paper we have investigated the mutual influence of drivers' trust and user experience in automated vehicles. We were interested in how subjective trust and UX correlate when modifying relevant parameters of both constructs (here: system performance of the AV, representing the most important criterium for appropriately calibrated trust, as well as usability and aesthetics of an IVIS as relevant UX parameters). Results of a driving simulator study, where 48 participants had to safely complete drives in an AV at level 2 while performing tasks on an IVIS, confirm that UX and trust influence each other and correlate. Participants were not able to solely adjust their trust levels to an objective measure of trustworthiness (system performance) as their judgment was strongly influenced by the UX of the IVIS (and vice-versa). Variations of investigated independent variables significantly affected both constructs of trust and UX. The study further confirms the existence of so-called "halo effects" in the context of AVs, which is an important finding as overtrust/overreliance already led to fatal accidents. Research investigating methods aiming to deal with trust issues should, thus, not only rely on subjective measurements of trust, but also consider and include user experience measures. Our study shows that level 2 driving may not be safely possible without making system performance accessible to drivers. Otherwise, the influence of design features could hinder drivers' ability to judge the trustworthiness of automated vehicles with the necessary objectivity.

ACKNOWLEDGMENTS

We applied the SDC approach for the sequence of authors. This work is supported under the FH-Impuls program of the German Federal Ministry of Education and Research, Grant Number 13FH7I01IA (SAFIR).

REFERENCES

[1] Lisanne Bainbridge. 1983. Ironies of automation. In *Analysis, Design and Evaluation of Man–Machine Systems 1982*. Elsevier, 129–135.

[2] Hanna Bellem, Barbara Thiel, Michael Schrauf, and Josef F Krems. 2018. Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation research part F: traffic psychology and behaviour* 55 (2018), 90–100.

[3] Johannes Beller, Matthias Heesen, and Mark Vollrath. 2013. Improving the driver–automation interaction: An approach using automation uncertainty. *Human factors* 55, 6 (2013), 1130–1141.

[4] Mathias Benedek and Christian Kaernbach. [n. d.]. Ledalab Software. http://www.ledalab.de/

[5] Mathias Benedek and Christian Kaernbach. 2010. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190, 1 (2010), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028 arXiv:1606.01836

[6] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.

[7] Wolfram Boucsein, Don C. Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T. Roth, Michael E. Dawson, and Diane L. Filion. 2012. Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 8 (2012), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

[8] JJ Braithwaite, DG Watson, Jones Robert, and Rowe Mickey. 2013. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. … (2013), 1–42. https://doi.org/10.1017.S0142716405050034

[9] K. A. Brookhuis, D. DE Waard, and S. H. Fairclough. 2003. Criteria for driver impairment. *Ergonomics* 46, 5 (apr 2003), 433–445. https://doi.org/10.1080/001401302/1000039556

[10] James D Brown and Warren J Huffman. 1972. Psychophysiological measures of drivers under actual driving conditions. *Journal of Safety Research* 4, 4 (1972), 172–178.

[11] C Collet, A Clarion, M Morel, A Chapon, and C Petit. 2009. Physiological and behavioural changes associated to the management of secondary tasks while driving. *Applied Ergonomics* 40, 6 (2009), 1041–1046. https://doi.org/10.1016/j.apergo.2009.01.007

[12] SAE On-Road Automated Vehicle Standards Committee et al. 2014. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE International* (2014).

[13] SAE On-Road Automated Vehicle Standards Committee et al. 2016. Human Factors Definitions for Automated Driving and Related Research Topics. *SAE International* (2016).

[14] Pieter Desmet and Paul Hekkert. 2007. Framework of product experience. *International journal of design* 1, 1 (2007).

[15] Pieter MA Desmet. 2012. Faces of product pleasure: 25 positive emotions in human-product interactions. *International Journal of Design* (2012).

[16] David Alexander Dickie and Linda Ng Boyle. 2009. Drivers' understanding of adaptive cruise control limitations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 53. Sage Publications Sage CA: Los Angeles, CA, 1806–1810.

[17] Verena Distler, Carine Lallemand, and Thierry Bellet. 2018. Acceptability and Acceptance of Autonomous Mobility on Demand: The Impact of an Immersive Experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 612.

[18] Florian N Egger et al. 2001. Affective design of e-commerce user interfaces: How to maximise perceived trustworthiness. In *Proc. Intl. Conf. Affective Human Factors Design*. 317–324.

[19] Mica R Endsley. 2017. Autonomous driving systems: a preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making* 11, 3 (2017), 225–238.

[20] Anna-Katharina Frison, Laura Aigner, Philipp Wintersberger, and Andreas Riener. 2018. Who is Generation A? Investigating the Experience of Automated Driving for Different Age Groups. *AutomotiveUI' 18, September 23-25, 2018, Toronto, Canada* (2018), in Press.

[21] Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, and Clemens Schartmüller. 2017. Driving Hotzenplotz: A Hybrid Interface for Vehicle Control Aiming to Maximize Pleasure in Highway Driving. *AutomotiveUI' 17 Adjunct, September 24-27, 2017, Oldenburg, Germany* (2017), pp. 6. https://doi.org/10.1145/3122986.3123016

[22] Eva Ganglbauer, Johann Schrammel, Stephanie Deutsch, and Manfred Tscheligi. 2009. Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. In *Workshop on user experience evaluation methods in product development*. Citeseer.

[23] Marc Hassenzahl. 2003. The thing and I: understanding the relationship between user and product. In *Funology*. Springer, 31–42.

[24] Marc Hassenzahl. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-computer interaction* 19, 4 (2004), 319–349.

[25] Marc Hassenzahl. 2008. User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th Conference on l'Interaction Homme-Machine*. ACM, 11–15.

[26] Marc Hassenzahl. 2018. The thing and I: understanding the relationship between user and product. In *Funology 2*. Springer, 301–313.

[27] Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. 2010. Needs, affect, and interactive products–Facets of user experience. *Interacting with computers* 22, 5 (2010), 353–362.

[28] Marc Hassenzahl and Andrew Monk. 2010. The Inference of Perceived Usability From Beauty. *HumanâĂŞComputer Interaction* 25, 3 (2010), 235–260. https://doi.org/10.1080/07370024.2010.500139 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/07370024.2010.500139

[29] Marc Hassenzahl and Andrew Monk. 2010. The inference of perceived usability from beauty. *Human–Computer Interaction* 25, 3 (2010), 235–260.

[30] Marc Hassenzahl, Annika Wiklund-Engblom, Anette Bengs, Susanne Hägglund, and Sarah Diefenbach. 2015. Experience-oriented and product-oriented evaluation: psychological need fulfillment, positive affect, and product perception. *International journal of human-computer interaction* 31, 8 (2015), 530–544.

[31] Renate Häuslschmid, Max von Buelow, Bastian Pfleging, and Andreas Butz. 2017. SupportingTrust in Autonomous Driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 319–329.

[32] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434.

[33] ISO 9241-11:1998(en) 1998. *Ergonomic requirements for office work with visual display terminals (VDTs) âĂŤ Part 11: Guidance on usability.* Standard. International Organization for Standardization, Geneva, CH.

[34] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71.

[35] Patrick W Jordan. 2003. Designing great stuff that people love. *From Usability to Enjoyment* (2003).

[36] Theresa T Kessler, Cintya Larios, Tiffani Walker, Valarie Yerdon, and PA Hancock. 2017. A Comparison of Trust Measures in Human–Robot Interaction Scenarios. In *Advances in Human Factors in Robots and Unmanned Systems*. Springer, 353–364.

[37] Jinwoo Kim and Jae Yun Moon. 1998. Designing towards emotional usability in customer interfacesâĂŤtrustworthiness of cyber-banking system interfaces. *Interacting with computers* 10, 1 (1998), 1–29.

[38] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, 4 (2015), 269–275.

[39] Johannes Maria Kraus, Jessica Sturn, Julian Elias Reiser, and Martin Baumann. 2015. Anthropomorphic agents, transparent automation and driver personality: towards an integrative multi-level model of determinants for effective driver-vehicle cooperation in highly automated vehicles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 8–13.

[40] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2017. Enhancing driving safety and user experience through unobtrusive and function-specific feedback. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. ACM, 183–189.

[41] H Landin. 2009. Anxiety and trust and other expressions of interaction [Doctoral dissertation]. *Göteborg, Sweden: Chalmers University of Technology* (2009).

[42] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 63–76.

[43] Effie Lai-Chong Law. 2011. The measurability and predictability of user experience. In *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 1–10.

[44] Effie Lai-Chong Law, Arnold POS Vermeeren, Marc Hassenzahl, and Mark Blythe. 2007. Towards a UX manifesto. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*. BCS Learning & Development Ltd., 205–206.

[45] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[46] Eva Lenz, Sarah Diefenbach, and Marc Hassenzahl. 2014. Aesthetics of interaction: a literature synthesis. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM, 628–637.

[47] Yung-Ming Li and Yung-Shao Yeh. 2010. Increasing trust in mobile commerce through design aesthetics. *Computers in Human Behavior* 26, 4 (2010), 673–684.

[48] Tao Lin, Masaki Omata, Wanhua Hu, and Atsumi Imamiya. 2005. Do physiological data relate to traditional usability indexes? *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future* (2005), 1–10. http://portal.acm.org/citation.cfm?id=1108368.1108405

[49] Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 1 (2011), 1.

[50] Todd Litman. 2017. *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada.

[51] Todd Litman. 2018. *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada.

[52] Andreas Löcken, Wilko Heuten, and Susanne Boll. 2015. Supporting lane change decisions with ambient light. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 204–211.

[53] Andrew Mackinnon, Anthony F Jorm, Helen Christensen, Ailsa E Korten, Patricia A Jacomb, and Bryan Rodgers. 1999. A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual differences* 27, 3 (1999), 405–416.

[54] Alexander Meschtscherjakov, Christine Döttlinger, Christina Rödel, and Manfred Tscheligi. 2015. ChaseLight: ambient LED stripes to control driving speed. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 212–219.

[55] Michael Minge and Manfred Thüring. 2018. Hedonic and pragmatic halo effects at early stages of user experience. *International Journal of Human-Computer Studies* 109 (2018), 13–25.

[56] Drew M Morris, Jason M Erno, and June J Pilcher. 2017. Electrodermal Response and Automation Trust during Simulated Self-Driving Car Use. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, 1759–1762.

[57] Bonnie M Muir. 1987. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies* 27, 5-6 (1987), 527–539.

[58] Brittany E Noah, Philipp Wintersberger, Alexander G Mirnig, Shailie Thakkar, Fei Yan, Thomas M Gable, Johannes Kraus, and Roderick McCall. 2017. First Workshop on Trust in the Age of Automated Driving. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. ACM, 15–21.

[59] Donald A Norman. 2004. Introduction to this special section on beauty, goodness, and usability. *Human-Computer Interaction* 19, 4 (2004), 311–318.

[60] Ingrid Pettersson, Florian Lachner, Anna-Katharina Frison, Andreas Riener, and Andreas Butz. 2018. A Bermuda Triangle?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 461.

[61] Bastian Pfleging, Maurice Rang, and Nora Broy. 2016. Investigating user needs for non-driving-related activities during automated driving. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 91–99.

[62] Christina Rödel, Susanne Stadler, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 1–8.

[63] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *IJIMAI* 4, 4 (2017), 40–44.

[64] Kennon M Sheldon, Andrew J Elliot, Youngmee Kim, and Tim Kasser. 2001. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology* 80, 2 (2001), 325.

[65] Kennon M Sheldon, Richard Ryan, and Harry T Reis. 1996. What makes for a good day? Competence and autonomy in the day and in the person. *Personality and social psychology bulletin* 22, 12 (1996), 1270–1279.

[66] David Sward and Gavin Macarthur. 2007. Making user experience a business strategy. In *E. Law et al.(eds.), Proceedings of the Workshop on Towards a UX Manifesto*, Vol. 3. 35–40.

[67] Noam Tractinsky, Adi S Katz, and Dror Ikar. 2000. What is beautiful is usable. *Interacting with computers* 13, 2 (2000), 127–145.

[68] Alexandre N Tuch and Kasper Hornbæk. 2015. Does Herzberg's notion of hygienes and motivators apply to user experience? *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 4 (2015), 16.

[69] Alexandre N Tuch, Sandra P Roth, Kasper HornbæK, Klaus Opwis, and Javier A Bargas-Avila. 2012. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior* 28, 5 (2012), 1596–1607.

[70] Heli Väätäjä, Tiina Koponen, and Virpi Roto. 2009. Developing practical tools for user experience evaluation: a case from mobile news

journalism. In *European Conference on Cognitive Ergonomics: Designing beyond the Product—Understanding Activity and User Experience in Ubiquitous Environments*. VTT Technical Research Centre of Finland, 23.

[71] Paul van Schaik, Marc Hassenzahl, and Jonathan Ling. 2012. User-Experience from an Inference Perspective. *ACM Trans. Comput.-Hum. Interact.* 19, 2, Article 11 (July 2012), 25 pages. https://doi.org/10.1145/2240156.2240159

[72] Trent W Victor, Emma Tivesten, Pär Gustavsson, Joel Johansson, Fredrik Sangberg, and Mikael Ljung Aust. 2018. Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel. *Human factors* (2018), 0018720818788164.

[73] Alan R Wagner, Jason Borenstein, and Ayanna Howard. 2018. Overtrust in the robotic age. *Commun. ACM* 61, 9 (2018), 22–24.

[74] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[75] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52 (2014), 113–117.

[76] Philipp Wintersberger and Andreas Riener. 2016. Trust in technology as a safety aspect in highly automated driving. *i-com* 15, 3 (2016), 297–310. https://doi.org/doi:10.1515/icom-2016-0034

[77] Philipp Wintersberger, Tamara von Sawitzky, Anna-Katharina Frison, and Andreas Riener. 2017. Traffic Augmentation as a Means to Increase Trust in Automated Driving Systems. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*. ACM, 17.

[78] Dennis Wixon. 2011. Measuring fun, trust, confidence, and other ethereal constructs: it isn't that hard. *interactions* 18, 6 (2011), 74–77.

[79] Peter Wright, John McCarthy, and Lisa Meekison. 2003. Making sense of experience. In *Funology*. Springer, 43–53.