Overtrust in External Cues of Automated Vehicles: An Experimental Investigation

Kai Holländer
LMU Munich, Germany
kai.hollaender@ifi.lmu.de

Philipp Wintersberger
CARISSMA, Technische Hochschule Ingolstadt, Germany
philipp.wintersberger@carissma.eu

Andreas Butz
LMU Munich, Germany
andreas.butz@ifi.lmu.de

ABSTRACT

The intentions of an automated vehicle are hard to spot in the absence of eye contact with a driver or other established means of communication. External car displays have been proposed as a solution, but what if they malfunction or display misleading information? How will this influence pedestrians' trust in the vehicle? To investigate these questions, we conducted a between-subjects study in Virtual Reality (N = 18) in which one group was exposed to erroneous displays. Our results show that participants already started with a very high degree of trust. Incorrectly communicated information led to a strong decline in trust and perceived safety, but both recovered very quickly. This was also reflected in participants' road crossing behavior. We found that malfunctions of an external car display motivate users to ignore it and thereby aggravate the effects of overtrust. Therefore, we argue that the design of external communication should avoid misleading information and at the same time prevent the development of overtrust by design.

CCS CONCEPTS

• **Human-centered computing** → *HCI theory, concepts and models*; *Virtual reality*; *Information visualization.*

KEYWORDS

automated driving, external car displays, pedestrian-vehicle-interaction, trust, user study

ACM Reference Format:
Kai Holländer, Philipp Wintersberger, and Andreas Butz. 2019. Overtrust in External Cues of Automated Vehicles: An Experimental Investigation. In *11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19), September 21–25, 2019, Utrecht, Netherlands.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3342197.3344528

AutomotiveUI '19, September 21–25, 2019, Utrecht, Netherlands

ACM ISBN 978-1-4503-6884-1/19/09…$15.00
https://doi.org/10.1145/3342197.3344528

Figure 1: VR simulation of an automated vehicle using the communication concept proposed by Fridman et al. [9].

1 INTRODUCTION

Automated vehicles (AVs) promise various advantages, such as improved safety, traffic flow, comfort, or mobility for new target groups. However, existing automated driving systems, mainly operating at SAE (Society of Automotive Engineers) level 2 [36], already led to life-threatening and even fatal accidents [43]. Hence, the public opinion about automated driving systems is diverse, and trust in automation could become a key issue for a potential success of automated vehicle technology [19]. Research on the interaction between humans and automated vehicles recently gained attention, and still remains an open challenge [5].

Prior research focused on a variety of strategies to communicate the car's current state or intention to vulnerable road users (VRUs, e.g., pedestrians or cyclists) [25]. VRUs then have to trust this information to avoid making mistakes when, e.g., crossing the road. At the same time, VRUs should not neglect the risk of deficient system actions, or underestimate the consequences of potential technology errors.

Various research institutions in the automotive industry develop concepts for highly automated vehicles. Examples include the Mercedes F 015 concept car[1], which projects a crosswalk on the road to signal pedestrians that they can

[1]The Mercedes-Benz F 015 Luxury in Motion; *last accessed: April 2019*

safely cross or the Semcon Smiling Car[2], which shows a smile on the front grille to indicate pedestrians if it is safe to cross the street. During testing and development, such systems might still have weaknesses in sensing or processing. However, the main cause for accidents remains the human factor[3]. Inadequate trust (or more precisely, overtrust) is suspected to be a main cause of the accidents that already happened with automated vehicles [40, 44]. Examples include a fatal crash with an activated Tesla Autopilot in 2016, or the Uber self-driving Taxi in 2018, where overreliance is suspected to play a major role. Thus, we raise the (research) question if similar situations could also occur within AVs and VRUs.

RQ: How does contradicting presentation of the intentions of an automated vehicle via an external car display influence other road users' trust and behavior?

More precisely, we investigated the influence of a malfunctioning external vehicle display on pedestrians' perceived safety in crossing scenarios, on their trust in the external car display and their confidence in the automated vehicle (AV). Malfunctions are known to impact trust and reliance behavior [20]. For example, Itoh et al. [15] investigated different occurrence patterns of errors and stated, that single malfunctions, although having an effect on trust, quickly recover under normal conditions. This might be relevant in the domain of automated driving, where even single errors can have drastically impacts. Thus, we conducted a user study (N = 18) in virtual reality, where pedestrians were encouraged to cross the road in front of an automated car. Like in similar studies in this domain, the AV communicated its yielding behavior (i.e., whether it yields the right of way and the VRU is allowed to cross the road) on an external display. For one group of participants, the displayed message always matched the actual behavior of the vehicle, while for the other group the message and the behavior conflicted. In these cases, the AV communicated that it would yield the right of way, but did not stop (or vice-versa) in one out of twelve trials. As of April 2019 there is, to the best of our knowledge, no published experiment explicitly addressing overtrust in the context of AV and VRU interaction.

Our results show that pedestrians do consider external car displays when crossing in front of a vehicle. A mismatch between displayed intentions and vehicle behavior motivated participants to wait unnecessarily in front of a stopping vehicle. Furthermore, when a malfunction occurred, perceived safety while crossing and confidence in the vehicle significantly decreased. Surprisingly, both recovered quickly. These results clearly indicate that further investigations regarding overtrust of pedestrians in automated vehicles is needed. We contribute some first insights on this issue, which might serve as a basis for developing future interaction concepts and studies. In particular, considering the issue of overtrust during the design of external cues is of high importance for the acceptance and safety of automated driving.

[2]The Smiling Car; *last accessed: April 2019*

[3]People and Autonomous Vehicle Accidents; *last accessed: April 2019*

2 RELATED WORK

This section presents related work in the context of (over-) trust in automation, trust in automated vehicles and insights regarding AV/VRU communication.

Trust in Automation

Overtrust in the context of human-computer interaction is understood as a false estimation of the risk while interacting with a machine. According to Wagner et al. [40], it includes two patterns or a combination of both: first, users underestimate the consequences if the system fails. Second, users underestimate the likelihood that a system will make serious mistakes at all. Norman [29] argued, that many accidents in cooperation with automated systems do rather result from inappropriate feedback than human error, and Parasuraman [30] shaped the terms use, disuse, and misuse: *Use* reflects proper system interaction, *disuse* (potentially as a result of distrust) prevents automation usage, and *misuse* (potentially emerging from overtrust) means to use a given system under the wrong circumstances. Muir [26] argued, that automated systems must provide well-designed decision aids that prevent both distrust and overtrust, with the goal to match operators' trust levels to an objective measure of trustworthiness ("calibration of trust"). One of the most influencing papers in the domain of trust in automation is the work of Lee and See [20]. In their work, they intensively discussed the impact of trust on reliance behavior, and further integrated former studies into a descriptive model. The model describes the relationship between capabilities of the automation and trust using three relevant factors – calibration (the degree to which trust matches system capabilities), resolution (the range-mapping of automation capabilities and trust), and specificity, that refers *"to the degree to which trust is associated with a particular component or aspect of the trustee"* [20]. If trust matches system capabilities, trust is calibrated. In contrast, overtrust means that the operator's trust exceeds the capabilities of the automation, and distrust describes a situation where trust is below objective automation performance [20]. Additional theoretical considerations have been discussed by Hoff and Bashir [12], who proposed a three-layered framework that distinguishes between dispositional (personality traits influencing trust already before system interaction), situational (contextual impact of the environment and internal characteristics of the operator), and learned (emerging from experience with system interaction) trust. They also state the strength of the relationship between trust and reliance being influenced by

the complexity of the automation, the novelty of a situation, the ability to compare automated with manual performance, and the operator's degree of decisional freedom [12].

Trust in Automated Vehicles

Trust in AVs has recently become a highly discussed topic among the AutomotiveUI community [28, 43], and there seems to be a consensus that trust in automation could become one of the major barriers that prevent a successful implementation of automated vehicle technology [19]. Inagaki and Itoh [14] presented a theoretical framework in the context of overtrust in advanced driver assistance systems (ADAS). The authors distinguish between overtrust and overreliance. According to them, overtrust describes an unrealistic assessment of the situation and leads to decisions appearing trustworthy even though they are not. Overreliance on an ADAS is a poor decision to an action that may result from overtrust (however, also from other influencing factors, such as situation awareness or workload [20]). For example, Tesla's "Autopilot" was involved in at least three fatal accidents within the last two years[4] due to an underestimation of the consequences and overreliance in the product. To maintain and calibrate trust in AVs, multiple strategies have been proposed, including (but not limited to) the provision of "why-and-how" information [16], so-called reliability/uncertainty displays [18, 27, 42], Augmented Reality aids [31] or anthropomorphic agents [17].

AV/VRU Communication

Pedestrians consider a variety of factors to decide whether it is safe to cross the road [37], such as vehicle speeds, size of safety gaps, vehicle movement, familiarity of the place, traffic density, etc. While some studies suggest that crossing decisions are mainly based on implicit interaction, such as perceived speed or gap size rather than explicit communication [1, 7, 47], recent experiments have shown that both – depending on the distance between the pedestrian and the vehicle – could be important [4]. Still, in automated driving the communication from a driver to a pedestrian becomes obsolete. Through automated vehicles, there will be an interaction triangle including on-board passengers, the vehicle's automation system and other traffic participants such as, pedestrians or cyclists [35]. In some situations a vehicle could even move unmanned e.g., to find a parking lot. Especially scenarios that are resolved with communication between road users (such as when a pedestrian crosses in front of a vehicle at an unregulated crossing), both entities should be able to avoid conflicts. Hence, when there is no human driver involved, the resulting communicational demands must be substituted by automated systems, whereas VRU/AV communication systems should help to increase safety and acceptance [11, 32, 33, 38, 46]. To that end, there are manifold concepts from research and industry to foster vehicle-to-pedestrian communication, e.g. tactile feedback via mobile devices [23], external car displays [1, 9, 13, 22, 23, 34], projections [3, 39] or physical attachments to the chassis [6, 23]. Although the importance of outward displays for crossing decisions is not completely clear, there is related work showing that external displays can significantly increase trust and confidence of pedestrians in such scenarios [13, 24].

[4]Tesla Autopilot Crashes and Causes; *last accessed: March 2019*

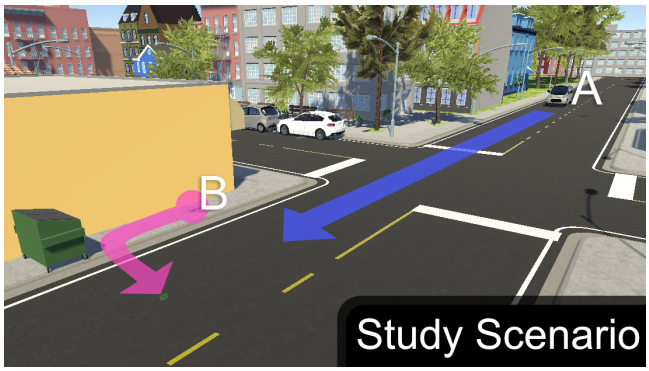

Figure 2: Simulated scenario with the movement paths of the vehicle (A) and the pedestrian (B).

In summary, overtrust in automation is a crucial issue especially in safety-critical environments, such as automated driving. When interacting with automated systems, there is a trade-off between disuse and misuse [30]. Existing research in the domain of automated vehicles primarily focuses on trust from the perspective of drivers and/or passengers. An increasing body of literature recently also addresses communication between AVs and VRUs, often focusing on visual appearances [5–7, 25]. In this context, we claim that there is a need to also consider trust and overtrust issues when designing AV-VRU interaction concepts.

3 RESEARCH APPROACH

Consequently, we raise the question, how VRUs react when being confronted with contradicting information on an external vehicle display. Furthermore, we want to investigate if there is a basis for overtrust. Since there is no related work on this issue in this context, we chose an exploratory approach. Therefore, we set up a study scenario which follows approaches from related work regarding pedestrian and AV interaction. The investigated scenario includes a straight road (see Figure 2), a pedestrian that intends to cross (represented by study participants) and an AV with an external display attached to the front grille, see Figure 1. The implemented display concept is inspired by Fridman et al. [9].

Their design adopts symbols known from US traffic lights: a green person and a yellow hand. In an online study (N = 200) that compared different approaches, this concept performed best [9]. Hence, we utilized a green person icon to signal pedestrians that the vehicle will come to a stop and yield. In contrast, the symbol of a yellow hand instructed pedestrians to wait. In this case the car continues driving without slowing down. As a novelty in research about external signals of AVs, we introduce the idea of malfunctions in the display concept and analyze human behavior and the development of trust in the resulting situation.

4 USER STUDY

A VR setup allowed us to conduct the experiment under laboratory conditions, while not creating any danger for participants. The study was conducted in accordance with the current version of the Helsinki Declaration[5]. We implemented an urban environment with a crossroad, an approaching automated vehicle, and no other moving traffic besides the participant representing the pedestrian. The automated vehicle resembles a Citroën C-Zero[6]. According to the manufacturer, this vehical provides a futuristic appearance and is especially designed for urban environments.

Study Design

The study consists of a between-subjects design with correct (matching) or incorrect (mismatching) display information as the independent variable. A total of 18 participants conducted 12 study trials each, resulting in 216 individual data points. Participants of the second group (g2) were always exposed to correct display information (match). For the first group (g1), a single malfunction appeared in the ninth of twelve trials (mismatch). This means that also participants in the "incorrect display" group could experience how the system is supposed to work and build up trust. During the intended display error, either the vehicle stopped although a yellow hand was presented or it did not yield the way while indicating the green person symbol. The cases alternated after each participant in the second group during the ninth trial. The frequency of the two matching display information (green person and vehicle stop, yellow hand and vehicle drive) was distributed evenly (counterbalanced) over all trials and participants (for both groups).

Participants

The study involved 18 individuals aged between 18 and 80 years (M = 31.83 years, SD = 19.89 years; eleven women, six men, one other). The high average age is due to the fact that two of the participants were in their mid 50s and two were older than 79 years. We have not observed any obvious health or cognitive limitations amongst our participants and VR sickness did not occur. Students represented 61 % of the participants and 39 % covered various occupational groups. There were no special prerequisites required from the test subjects other than the ability to walk. About 27.8 % reported to walk for less than 30 min per day on average. A majority of 55.5 % stated to walk for more than 30 min and less than 60 min, whereas 16.7 % walk more than 60 min on a daily basis. All participants had normal or corrected to normal vision and were recruited via internal e-mail lists, social media channels and personal invitations. As a compensation, attendees received an online marketplace voucher worth five euros. The participants were distributed as evenly as possible to both groups, see Table 1. 'Risk taking' refers to self reported tendency to take risks during everyday road crossings by foot and 'av. walking time' represents reported time of daily walking. In order to distribute participants in equal groups and to assess their data more accurately, subjects to rated their individual tendency to take risks during crossing decisions at three levels: *low*, *medium* and *high*.

Table 1: Distribution of participants for groups 1 & 2.

	G1 (Mismatch)	**G2** (Match)
N	9	9
∅ age $_{\text{min \mid max}}$	30.55 years $_{18 \mid 80}$	32.13 years $_{19 \mid 79}$
SD age	19.80 years	19.47 years
∅ walking time	30 - 60 min	30 - 60 min
Gender	6f, 2m, 1o	5f, 4m
Risk taking	medium: 66.67% low: 33.33%	medium: 66.67% low: 33.33%

Task

Participants started on the sidewalk as shown in Figure 2. From there they were told to walk straight ahead and follow the pavement if possible. After about three meters a waste container occupied their path so that crossing the street became inevitable. As an additional motivation to step on the road we placed a banknote in the middle of the street (at the tip of the red arrow in Figure 2). When pedestrians reached the waste container, an automated vehicle appeared at a distance of 39 m with a constant speed of 30 km/h. Thus, the overall task for participants was to follow the sidewalk and then to cross the road. We did clarify that, as in real life, there is traffic on the street which should be considered.

Procedure

First, participants were welcomed by the examiner and introduced to the study task (see Section *Task*). Furthermore, we

[5]WMA Declaration of Helsinki; *last accessed: April 2019*
[6]Citroën C-Zero; *last accessed: April 2019*

explained the *green person* and *yellow hand* symbols. All participants signed a consent form and filled in a pre-study questionnaire concerning demographics. In addition, we asked for an individual estimation of the personal risk taking level in traffic situations. The starting position in the room was marked with tape on the ground. Participants were asked by the examiner to stand on this mark and put on the head-mounted display (HMD). Each participant completed twelve trials in Virtual Reality and hence, had to decide twelve times to cross or not to cross a road within the presented scenario. Between all twelve runs, subjects rated their perception of safety while crossing, their trust in the external car display and their confidence in the automated vehicle. In order to interrupt immersion as little as possible, participants kept the HMD on and received the questions via headphones orally. In addition, the experimenter recorded in writing whether the test persons hesitated before crossing. The vehicle behavior and the display were pseudo-randomly assigned, so that participants could not predict what the vehicle or the display would do. The HMD's headphones played a background noise typically heard in a city near a street to increase immersion. After twelve trials, respondents completed a final questionnaire including their personal perception of the car and its displays. Each participant spent $\approx$ 45 min in the lab.

Apparatus

The room dimensions are 8.6m by 3.6m with a physical movement area of approximately 3m by 3m. Walls in the real world were matched with walls, buildings or other objects in the virtual world, for example the yellow building in Figure 2. Hence, physical limitations of the real world were concealed by unobtrusive barriers in the virtual world. We used an HTC Vive (first generation) VR setup with a corresponding lighthouse tracking system. The simulation ran on a Windows 10 PC including an Nvidia GTX 1980Ti graphics card, an IntelCore i7- 6700k processor, and 16GB of RAM. The study environment was created in Unity 2018.2.0f2.

Measures

Pedestrian Behavior. Throughout the experiment we recorded three types of events from the simulation in a csv log file. Each event includes a Unix timestamp, the position oft the vehicle (x,y an z coordinates) and the position of the pedestrian (x,y and z coordinates). The first event marks the beginning of a study course (*"Beginning"*). For this event, only the timestamp matters, since start-coordinates were the same for each run. The second event is triggered if a pedestrian steps on the road in front of the vehicle (*"Stepped"*). The third event records collisions. However, a collision did not occur during this experiment. Furthermore, we calculated decision time if pedestrians decided to cross the street. The decision time (in seconds) is the difference between the timestamps *"Beginning"* and *"Stepped"*. We recorded with a binary option whether pedestrians decided to wait. Additionally, the examiner noted when hesitating behavior was observed. Hesitations were recorded if participants did not follow their decision consequently, but either slowed down, paused or changed their mind. For example, two participants moved towards the road, stopped briefly and then decided to wait.

Perceived Safety, Trust & Confidence. After each individual run, participants rated how *safe* they felt when the automated vehicle approached them. To that end, we presented a five-point Likert scale (1: very unsafe; 5: very safe). In addition, pedestrians stated after each run how much they trusted the external display of the automated vehicle and how *confident* they felt in the behavior of the vehicle. Both questions were again rated on a five-point Likert scale (1: very little trust / confidence; 5: very high trust / confidence). We specifically asked about the display and the vehicle to identify discrepancies for both entities in this context.

Interaction With the System. In the final questionnaire, we asked the following open questions: "Did you notice anything special about the car? If so, what?", "How did you feel about interacting with the vehicle and its displays?", and "Did the fact that the vehicle was automated affect your decision or behavior? If so, how?". In addition, the participants used a *yes* or *no* radio button to indicate whether the vehicle behaved as expected or not.

5 RESULTS

In this section, we first report our observations of pedestrian behavior, followed by the quantitative results regarding perceived safety, trust and confidence. Finally, we present qualitative insights from the post-study questionnaire regarding interaction with the system. In the text below, group one (g1) refers to participants who experienced a mismatch of displayed information and vehicle behavior in the ninth of twelve study runs. Participants in group two (g2) were exposed to consistently matching display content.

Results on Pedestrian Behavior

In a pre-study questionnaire participants reported to have a low (33.33%) or medium (66.67%) tendency to take risks in traffic. No one stated to have a high tendency for taking risks when crossing a road during everyday situations.

In all 207 runs, which did not include any malfunction of the display, participants crossed the road if they saw a green person on the display of the car. If a yellow hand was seen, all participants waited accordingly. Hence, if the displayed content was consistent with the behavior of the car, participants followed the indicated instructions. In the nine runs with an erroneous display, all subjects decided not to cross the road regardless of the displayed symbol and of

Table 2: Perceived safety in crossing decisions, trust in the display and confidence in the vehicle for both groups and in total.

	All (N = 216)			G1 (Mismatching Display; N = 108)			G2 (Matching Display; N = 108)		
	Safety	**Trust**	**Confidence**	**Safety**	**Trust**	**Confidence**	**Safety**	**Trust**	**Confidence**
Mean	4.09	4.29	3.79	4.31	4.44	4.01	3.88	4.13	3.56
SD	1.24	1.08	1.30	1.13	0.92	1.28	1.30	1.20	1.30
Median (min; max)	5 (1;5)	5 (1;5)	4 (1;5)	5 (1;5)	5 (1;5)	5 (1;5)	4.5 (1;5)	5 (1;5)	4 (1;5)

whether the car stopped or continued driving. Hence, in case of a malfunction all participants waited even if the vehicle came to a complete stop.

For both groups we saw a hesitating behavior. More than 50% of all hesitations *(22 in 216 runs)* occurred during the first three cycles. For group two (match), hesitating behavior decreased after the first three runs and formed an average of 10% over all 108 cycles. In trials nine to twelve 5.5% of participants in group two (match) hesitated. The subjects from group one (mismatch) hesitated in 13% of their 108 observations, mainly during the ninth run (77.8%). In trials nine to twelve 23.5% of participants in group one hesitated.

Decision time only varied at the beginning of the first three runs and then aligned between 9 and 16 seconds. We measured an average decision time of 13.25 seconds to cross the road. Looking at both groups individually, it can be seen that the group with a mismatching display took slightly longer to cross the road (g1 'mismatch': 13.68 seconds; g2 'match': 12.57 seconds). Group 2 included an outlier where one person waited for more than 30 seconds prior to crossing the street which was removed for the analysis of crossing times. From the ninth to the twelfth run, there were no significant differences in the decision times between both groups.

Table 2 shows descriptive results of perceived safety during the crossing decision, trust in the external car display and confidence in the vehicle. Kolmogorov-Smirnov comparisons show that the data from our independent samples is not normally distributed. Hence, we performed Man-Whittney-U (Wilcoxon rank-sum) tests for this evaluation.

Table 3 contains the resulting U-values (indicating how many ranking values of the other variables are lower overall), z-values (z distribution with critical value of 1.96), p-values and Pearson's correlation coefficient to assess the meaning of the p-values. As a result, both groups show significant ($\alpha = 0.05$) differences regarding perceived safety while crossing and the confidence in the vehicle. Calculated corresponding correlation effects for safety and confidence indicate a strong effect size ($r \geqslant 0.50$) [2]. Therefore, even a single malfunction of an external car display severely influences perceived safety and confidence in the interaction between AVs and VRUs.

The boxplot in Figure 3 illustrates the distribution of confidence in AVs for each group as an example. Surprisingly, trust

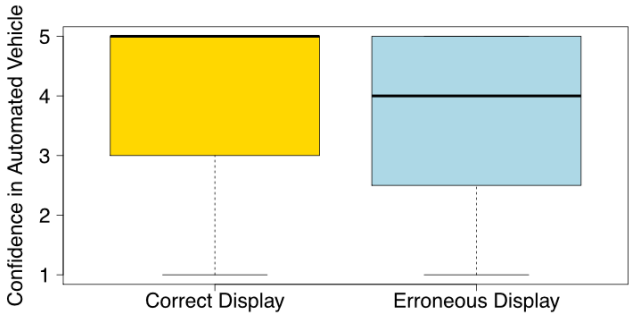

Figure 3: Perceived confidence in automated vehicle for both groups (yellow: g2 'match'; blue: g1 'mismatch').

in the external car display did not show a significant difference. However, there is a similar progress in the development of measured safety, trust and confidence, see Figure 4. Thus, initial values start at the upper half of the scale and increase slightly. For participants in the first group, the ninth run shows a strong decline for all three independent variables, which recovers in the tenth cycle.

Results of Post-Study Questionnaire

We asked participants if they noticed anything special about the car. The most stated answer was: "nothing" or "no" (66.7%). Two subjects stated that no driver could be seen and another two found the display on the car special, and described it as noteworthy. Participants reported their feelings when interacting with the automated vehicle and its display as "Safe and good" (38.9%), "Negatively biased, insecure, anxious" (22.2%), or "It is unfamiliar" (16.7%). Furthermore, P12 (g2)

Table 3: Results of a Man-Whittney-U test regarding perceived safety in crossing decisions, trust in the display and confidence in the vehicle with corresponding correlation coefficients (r) in accordance to Pearson.

	U	**z**	**p**	**r**
Safety	4814	-2.21	0.01	0.74
Trust	5149	-1.48	0.06	0.49
Confidence	4567	-2.75	< 0.01	0.91

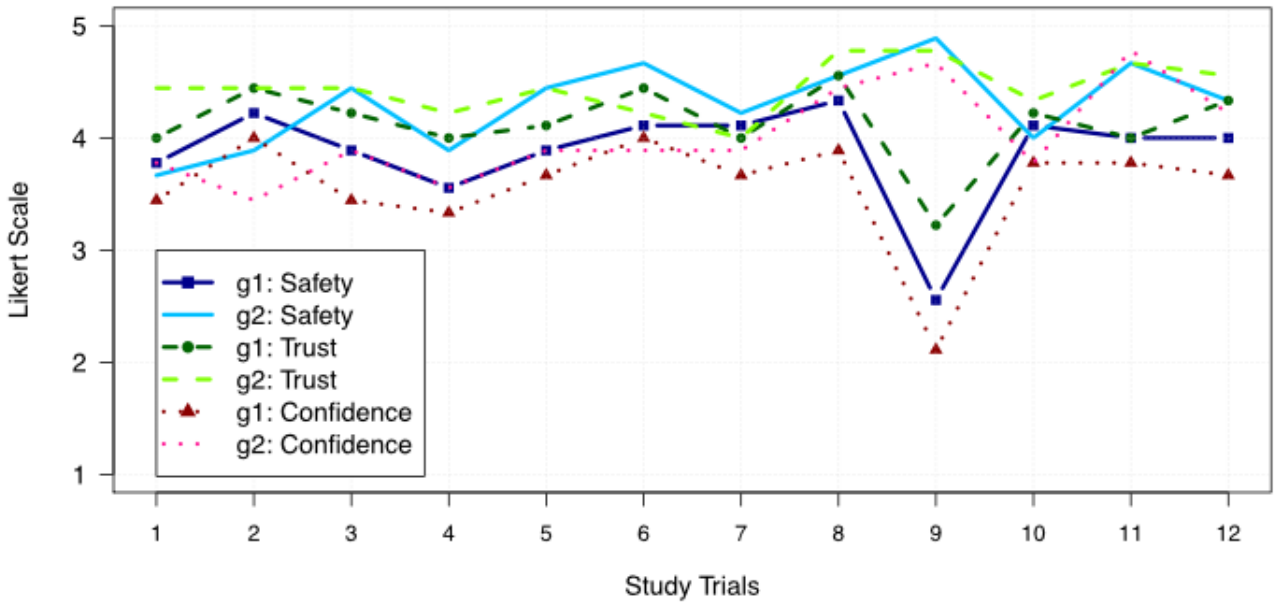

Figure 4: Mean values for perceived safety while crossing, trust in the externals display and confidence in the AV for both groups (g1: mismatching AV signals; g2: matching signals). Trials are on the horizontal axis, likert scale values on the vertical. In the ninth trial a mismatch in displayed information occurred for participants of group 1.

mentioned that "the display is helpful to determine if the car wants to let you by" and P17 (g1) noted: "I was more alert than usual". Answers to a question about whether the vehicle being automated influenced the decision or behavior of the participants were "no" (22.2%), "yes, I have less confidence in the car" (22.2%), and "I waited longer to see if the car would really stop" (22.2%). Additionally, two participants (one of each group) stated that the interaction felt rather unusual and that one has to learn to get used to the car. Finally, we asked if the car behaved as expected via a radio button. Most participants confirmed and clicked on "Yes" (89 %). Only two of 18 answered with "no", both were in g1 ('mismatch').

6 DISCUSSION

We expected trust levels to be low in the beginning and then to increase, as participants experience the scenario. However, initial trust in the external car display (ECD) and confidence in the automated vehicle were already high (see Figure 4). Trust in the ECD was always higher than confidence in the vehicle, but both developed similarly. Apparently participants trust a feature of a system (the display) more than the system as a whole. In contrast, a malfunction of the display directly influenced the perception of the complete vehicle. It seems that a failure of a subsystem communicates that the entire system is faulty. This finding is in line with results from Frison et al. [10]. Thus, when deploying poor external cues, the acceptance of the whole vehicle might be affected. This relates to low "functional specificity" according to the model of appropriate trust by Lee and See [20] (the degree to which users can distinguish between different subsystems of an automated system). These insights might not be limited to the scope of AVs and external cues, but are supposedly effective in other domains of human-computer interaction as well. For example, in the interplay of humans and robots or machines in households, health care or industrial factories.

The results further show that an erroneous display has a negative impact on perceived safety while crossing, trust in the ECD and confidence in the AV. Wrong information decreases perceived safety, trust and confidence, but only if actual malfunction appears. Surprisingly, these attributes recover directly afterwards. This could be interpreted as potentially high "temporal specificity" (changes in system capabilities are quickly reflected in trust levels [20]). However, it could also be an indicator for overtrust, especially when considering the results of the post-test questions, where only two 2 of 18 participants answered that the vehicle did not behave as they would have expected it to.

After the system failed for group two, hesitations in the following trials increased. However, there was no significant increase in decision times measured. Several error-free runs

strengthened the trust in the system. Trust can therefore be increased (or remains constant at a high level) when subjects continuously experience interaction with an automated vehicle as expected. This is in line with prior work, which indicated that single (or a small number of discrete) malfunctions do not sustainably impair trust (in comparison to continuous patterns of errors) [15]. However, Itoh et al. also concluded, that when subjects *"experience more individual malfunctions, they appear to become less sensitive to the malfunctioning"* [15]. Thus, it will be important to investigate different patterns in the future, as such habits could lead to potentially dangerous behavior in traffic.

Additionally, all subjects exposed to a correct display felt safe as the car approached them, and acted as indicated on the display. This feeling of safety also strengthened confidence in the vehicle and could develop into overtrust. Wagner and Koopman [41] state that people learn to inappropriately trust automated systems, and that they are not good at searching for errors. Users rather tend to assume that the system will do its job well. For safety-critical situations with irreversible outcome, this can lead to severe consequences.

As described above, 12 of 18 participants stated in the final questionnaire that they had not noticed anything special about the car, although nine participants were exposed to wrongly displayed instructions. For a majority of 89 % the vehicle did what they expected, although we had explained before the experiment how the symbols should correspond to vehicle movements. Nobody told the examiner that the vehicle or display had made a mistake. This is in line with Fitts' [8] findings, which include that people are not good at monitoring automated systems. It is probably more comfortable to trust a system and assume that it will do its job flawlessly than staying alert and questioning it.

Furthermore, all participants of the mismatch group decided not to cross during the ninth run. Surprisingly, even participants experiencing the vehicle coming to a stop waited. They told the examiner that they did not not want to cross because of the displayed "yellow hand", and reported that confidence in the display decreased drastically during this run. We can therefore claim that people actually do consider (even faulty) external car displays when taking a decision to cross. This is a valuable finding, since there are contrarily statements about the role of explicit and implicit communication in crossing decisions in the automotive research domain [1, 7**?**]. Our results indicate that ECDs could indeed become a valuable part of automated vehicles.

Key Aspects for Safe External Vehicle Cues

This study was based around a crossing scenario and took place in VR. Hence, participants were able to completely focus on the vehicle, without any distractions. In the real world, many decisions may be taken less consciously while walking as a pedestrian. We see that many different approaches to communicating with VRUs have been proposed [6]. Some of them go beyond binary information (walk/wait), and provide multiple different messages to be interpreted. Different vehicles with different forms of such communication could drastically increase the complexity from the perspective of VRUs, who have to interpret all the cues provided by vehicles in the vicinity. Especially scenarios with multiple vehicles (e.g., mixed traffic), multiple VRUs, (where pedestrians might be distracted, e.g., due to smartphone usage), might quickly become ambiguous and may lead to misconduct. This could become dangerous if concentration regarding the vehicles' movements decreases, and overtrust because of ECDs perceived as flawless develops in real life. In such situations, a single malfunction, or misinterpretation could lead to severe consequences. Also, misunderstood meanings of a display could motivate risky behavior. While some interpret an indication as a crossing instruction, Zhang et al. [45] show that others interpret external cues as intentions of the vehicle. In accordance to Wagner [40], we thus come to the conclusion that overtrust needs a holistic approach to be overcome. Especially, in the context of VRU/AV communication. Through the results of this study, we identified three key aspects in order to foster safe external vehicle cues:

(1) **Overtrust/overreliance should be recognized as an important aspect in the design process of external vehicle cues.** By considering trust-related issues in the first design iterations, concepts for reducing negative impact could be included in prototypes and be evaluated at an early stage.
(2) **Vulnerable road users need to be trained on how to cooperate with external cues of automated vehicles.** Especially, if the amount of automated vehicles on public roads increases, people should receive support on how to interpret external cues and always consider that technology may ultimately fail.
(3) **Developers should reach an agreement on how to communicate safety-related cues for future traffic.** Agreeing on a universal design language (similar to the appearance of most traffic signs, e.g., the stop sign) could help reducing mental overload during interactions and the complexity of various designs and therefore avoid misunderstandings.

Limitations

A limitation of this study is the small sample size, especially for a between-groups study design. Nevertheless, we collected 216 unique data points. Still, the results gained might not be generally accurate and should be interpreted as a trend indicator helping to identify relevant aspects for future

work. Another limiting aspect for a general validity is the usage of five-point Likert questions for perceived safety, trust and confidence. These attributes are complex and could each be evaluated with a validated questionnaire. However, since this work presents a first endeavor to investigate overtrust in external car displays at all, we believe that finding reasons to investigate the topic further is already a first valuable insight. Additional aspects for further research can also be found in the literature. For example, Lewis et al. [21] aim to uncover the role of trust in human-robot interaction. They point out that user studies investigating trust in this context often lack a definition of trust for participants. Therefore, individual participants within the same study might perceive trust differently. Hence, the comparability of studies regarding (over)trust in human-computer interaction is generally questionable. Nevertheless, an identification of overtrust through behavioral observations and interviews with participants can uncover potential issues.

Virtual Reality might also influence pedestrians' behavior. For example, it was not possible to cross the entire street in our simulation because the HTC Vive Headset can only detect a diagonal range of five meters. Therefore, participants were informed that it was sufficient to walk a few steps if a decision to cross was made. This unsettled some of the participants as they were afraid to run against the opposite wall. Three participants stated that this was very unusual in comparison to their behavior in the real world. Hence, it can be questioned whether the results of this work can be directly transferred to the real world. The high initial trust level and fast trust recovery might also be influenced by the VR setting. However, Virtual Reality enabled us to provide a controlled environment without the influence of e.g., daylight, other road users or weather. Additionally, it also ensured that no participant could be harmed in case of a collision. This factor is especially important for our scenario, since false information on the external car display could have provoked an accident. On the other hand, participants might also be aware that there is no physical harm to expect and therefore behave more risky.

7 CONCLUSION & FUTURE WORK

This study presents a first experiment on overtrust in external cues of automated vehicles. We investigated how presenting contradicting intentions of an automated vehicle via an external car display influences pedestrians' trust and behavior. Our results suggest that a single malfunction on an external car display influences pedestrian behavior and the perception of the automated system significantly, in the situation when the malfunction occurs. Nevertheless, reduced perceived safety, confidence and trust recover quickly. Therefore, a basis for overtrust is identified. Additionally, we present three suggestions on how to overcome overtrust in pedestrian-to-vehicle communication.

The main insight from this study is that further research in the domain of trust in automation from the perspective of VRUs (e.g., pedestrians and cyclists) is needed. This is not only relevant for safety, but also in terms of acceptance. For example, one participant stated: *"without a driver I can't find a car safe"*. Therefore, a well-thought interaction design for vehicle-to-VRU communication is necessary to foster safety, acceptance and thereby the overall success of AVs.

Future work should also include physiological measures to gather objective and unbiased data. Additionally, there are many possibilities to extend this initial study, such as more participants, study trials and misbehaviors. Investigating the long term effects of overtrust in external car displays and automated vehicles in general remains an open challenge.

ACKNOWLEDGMENTS

We thank Bao Loi Quach for her commitment in implementing this study. This work is supported under the FH-Impuls program of the German Federal Ministry of Education and Research (BMBF) under Grant Number 13FH7I01IA (SAFIR).

REFERENCES

[1] Michael Clamann, Miles Aubert, and Mary L Cummings. 2017. *Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles.* Technical Report. National Science Foundation.

[2] Jacob Cohen. 1992. A Power Primer. *Psychological Bulletin* 112, 1 (1992), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

[3] Ashley Colley, Jonna Häkkilä, Meri-Tuulia Forsman, Bastian Pfleging, and Florian Alt. 2018. Car Exterior Surface Displays: Exploration in a Real-World Context. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis '18).* ACM, New York, NY, USA, Article 7, 8 pages. https://doi.org/10.1145/3205873.3205880

[4] Debargha Dey, Walker Francesco, Martens Marieke, and Terken Jacques. 2019. Gaze Patterns in Pedestrian Interaction with Vehicles - Towards Effective Design of External Human-Machine Interfaces for Automated Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19).* ACM, New York, NY, USA.

[5] Debargha Dey, Azra Habibovic, Maria Klingegård, Victor Malmsten Lundgren, Jonas Andersson, and Anna Schieben. 2018. Workshop on Methodology: Evaluating Interactions between Automated Vehicles and Other Road Users—What Works in Practice?. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* ACM, 17–22.

[6] Debargha Dey, Marieke Martens, Chao Wang, Felix Ros, and Jacques Terken. 2018. Interface concepts for intent communication from autonomous vehicles to vulnerable road users. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* ACM, 82–86.

[7] D. Dey and J.M.B. Terken. 2017. Pedestrian interaction with vehicles: roles of explicit and implicit communication. In *AutomotiveUI '17 Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 24-27 September 2017, Oldenbourg, Germany.* Association for Computing Machinery, Inc, United States, 109–113. https://doi.org/10.1145/3122986.3123009

[8] Paul M Fitts. 1951. Human engineering for an effective air-navigation and traffic-control system. (1951).

[9] Lex Fridman, Bruce Mehler, Lei Xia, Yangyang Yang, Laura Yvonne Facusse, and Bryan Reimer. 2017. To Walk or Not to Walk: Crowdsourced Assessment of External Vehicle-to-Pedestrian Displays. *CoRR* abs/1707.02698 (2017). arXiv:1707.02698 http://arxiv.org/abs/1707.02698

[10] 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 *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 144, 13 pages. https://doi.org/10.1145/3290605.3300374

[11] Nicolas Guéguen, Sebastien Meineri, and Chloé Eyssartier. 2015. A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety Science* 75 (06 2015). https://doi.org/10.1016/j.ssci.2015.01.018

[12] 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.

[13] Kai Holländer, Ashley Colley, Christian Mai, Jonna Häkkilä, Florian Alt, and Bastian Pfleging. 2019. Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2019)*. ACM, New York, NY, USA, 11. https://doi.org/10.1145/3338286.3340138

[14] Toshiyuki Inagaki and Makoto Itoh. 2013. Human's Overtrust in and Overreliance on Advanced Driver Assistance Systems: A Theoretical Framework. *International Journal of Vehicular Technology* 2013 (2013), 1–8. https://doi.org/10.1155/2013/951762

[15] Makoto Itoh, Genya Abe, and Kenji Tanaka. 1999. Trust in and use of automation: their dependence on occurrence patterns of malfunctions. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, Vol. 3. IEEE, 715–720.

[16] 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.

[17] 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.

[18] 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.

[19] John D Lee and Kristin Kolodge. 2018. Understanding attitudes towards self-driving vehicles: Quantitative analysis of qualitative data. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62. SAGE Publications Sage CA: Los Angeles, CA, 1399–1403.

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

[21] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. *The Role of Trust in Human-Robot Interaction*. Springer International Publishing, Cham, 135–159. https://doi.org/10.1007/978-3-319-64816-3_8

[22] Yeti Li, Murat Dikmen, Thana G Hussein, Yahui Wang, and Catherine Burns. 2018. To Cross or Not to Cross: Urgency-Based External Warning Displays on Autonomous Vehicles to Improve Pedestrian Crossing Safety. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 188–197.

[23] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 429, 12 pages. https://doi.org/10.1145/3173574.3174003

[24] Milecia Matthews, Girish Chowdhary, and Emily Kieson. 2017. Intent Communication between Autonomous Vehicles and Pedestrians. *CoRR* abs/1708.07123 (2017). arXiv:1708.07123 http://arxiv.org/abs/1708.07123

[25] Alexander G Mirnig, Philipp Wintersberger, Alexander Meschtscherjakov, Andreas Riener, and Susanne Boll. 2018. Workshop on Communication between Automated Vehicles and Vulnerable Road Users. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 65–71.

[26] 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.

[27] Brittany E. Noah, Thomas M. Gable, Shao-Yu Chen, Shruti Singh, and Bruce N. Walker. 2017. Development and Preliminary Evaluation of Reliability Displays for Automated Lane Keeping. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17)*. ACM, New York, NY, USA, 202–208. https://doi.org/10.1145/3122986.3123007

[28] 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 (AutomotiveUI '17)*. ACM, New York, NY, USA, 15–21. https://doi.org/10.1145/3131726.3131733

[29] Donald A Norman. 1990. The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 327, 1241 (1990), 585–593.

[30] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors* 39, 2 (1997), 230–253.

[31] Wintersberger Philipp, Frison Anna-Katharina, Andreas Riener, and Tamara von Sawitzky. 2019. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *Presence-Teleoperators and Virtual Environments* (2019), 27–1.

[32] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. 2017. Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. 264–269. https://doi.org/10.1109/IVS.2017.7995730

[33] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. 2018. Towards Social Autonomous Vehicles: Understanding Pedestrian-Driver Interactions. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. 729–734. https://doi.org/10.1109/ITSC.2018.8569324

[34] Amir Rasouli and John K Tsotsos. 2018. Autonomous Vehicles that Interact with Pedestrians: A Survey of Theory and Practice. *arXiv preprint arXiv:1805.11773* (2018).

[35] Anna Schieben, Marc Wilbrink, Carmen Kettwich, Ruth Madigan, Tyron Louw, and Natasha Merat. 2018. Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations. *Cognition, Technology & Work* (15 Sep 2018). https://doi.org/10.1007/s10111-018-0521-z

[36] SAE Society of Automotive Engineers. 2018. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016 Ground Vehicle Standard).* https://doi.org/10.4271/J3016_201806

[37] Matúš Šucha. 2014. Road users' strategies and communication: driver-pedestrian interaction. *Transport Research Arena (TRA)* (2014).

[38] Matus Sucha, Daniel Dostal, and Ralf Risser. 2017. Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention* 102 (2017), 41 – 50. https://doi.org/10.1016/j.aap.2017.02.018

[39] Nguyen Trung Thanh, Holländer Kai, Hoggenmueller Marius, Parker Callum, and Tomitsch Martin. 2019. Designing for Projection-based Communication between Autonomous Vehicles and Pedestrians. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19).* ACM, New York, NY, USA. https://doi.org/10.1145/3342197.3344543

[40] Alan R. Wagner, Jason Borenstein, and Ayanna Howard. 2018. Overtrust in the Robotic Age. *Commun. ACM* 61, 9 (Aug. 2018), 22–24. https://doi.org/10.1145/3241365

[41] Michael Wagner and Philip Koopman. 2015. A Philosophy for Developing Trust in Self-driving Cars. In *Road Vehicle Automation 2*, Gereon Meyer and Sven Beiker (Eds.). Springer International Publishing, Cham, 163–171.

[42] Philipp Wintersberger, Dmitrijs Dmitrenko, Clemens Schartmüller, Anna-Katharina Frison, Emanuela Maggioni, Marianna Obrist, and Andreas Riener. 2019. S(C)ENTINEL: Monitoring Automated Vehicles with Olfactory Reliability Displays. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19).* ACM, New York, NY, USA, 538–546. https://doi.org/10.1145/3301275.3302332

[43] Philipp Wintersberger, Brittany E. Noah, Johannes Kraus, Roderick McCall, Alexander G. Mirnig, Alexander Kunze, Shailie Thakkar, and Bruce N. Walker. 2018. Second Workshop on Trust in the Age of Automated Driving. In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18).* ACM, New York, NY, USA, 56–64. https://doi.org/10.1145/3239092.3239099

[44] Philipp Wintersberger and Andreas Riener. 2016. Trust in technology as a safety aspect in highly automated driving. *i-com* 15, 3 (2016), 297–310.

[45] Jingyi Zhang, Erik Vinkhuyzen, and Melissa Cefkin. 2017. Evaluation of an autonomous vehicle external communication system concept: a survey study. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 650–661.

[46] Xiangling Zhuang and Changxu Wu. 2014. Pedestrian gestures increase driver yielding at uncontrolled mid-block road crossings. *Accident Analysis & Prevention* 70 (2014), 235 – 244. https://doi.org/10.1016/j.aap.2013.12.015

[47] Raphael Zimmermann and Reto Wettach. 2017. First Step into Visceral Interaction with Autonomous Vehicles. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17).* ACM, New York, NY, USA, 58–64. https://doi.org/10.1145/3122986.3122988