Talking Cars, Doubtful Users—A Population Study in Virtual Reality

Shadi Derakhshan, Farbod Nosrat Nezami, Maximilian Alexander Wächter, Artur Czeszumski, Ashima Keshava, Hristofor Lukanov, Marc Vidal De Palol, Gordon Pipa, and Peter König

***Abstract*—Autonomous vehicles represent a significant development in our society, and their acceptance will largely depend on trust. This study investigates strategies to increase trust and acceptance by making the cars' decisions. For this purpose, we created a virtual reality (VR) experiment with a self-explaining autonomous car, providing participants with verbal cues about crucial traffic decisions. First, we investigated attitudes toward self-driving cars among 7850 participants using a simplified version of the Technology Acceptance Model (TAM) questionnaire. Results revealed that female participants are less accepting than male participants, and that there is a general decline among all genders. Otherwise in general, a self-explaining car has a positive impact on trust and perceived usefulness. Surprisingly, it adversely affected the intention to use and perceived ease of use. This entails dissociation of trust from the other items of the questionnaire. Second, we analyzed behavioral of 26 750 participants to investigate the effect of self-explaining systems on head movements during the VR drive. We observed significant differences in head movements during the critical events and the baseline periods of the drive between the three driving conditions. Additionally, we demonstrated positive correlations between head movement parameters and the TAM scores, where trust showed the lowest correlation. This provides further evidence of the dissociation of trust from other TAM items. These findings illustrate the benefits of combining subjective questionnaire data with objective behavioral data. Overall, the outcomes indicate a partial dissociation of self-reported trust from intention to use and objective behavioral data.**

***Index Terms*—Autonomous vehicles, demographic differences, head movement, technology acceptance, virtual reality (VR).**

Manuscript received July 27, 2021; revised December 15, 2021 and February 10, 2022; accepted March 11, 2022. Date of publication May 25, 2022; date of current version July 14, 2022. This work was supported in part by the University of Osnabrück in cooperation with graduate college "Vertrauen und Akzeptanz in erweiterten und virtuellen Arbeitswelten" (FN), in part by German Research Foundation DFG, which supported this project in the context of funding the Research Training Group "Situated Cognition" under Grant GRK 2185/1.(SD), in part by the German Federal Ministry of Education and Research for the Project ErgoVR (Entwicklung eines Ergonomie-Analyse-Tools in der virtuellen Realität zur Planung von Arbeitsplätzen in der industriellen Fertigung) under Grant 16SV8052(AK), in part by NBank cooperation Project "Smart Fidelity Interaktionssystem zur Steigerung der Realitätsnähe bei der Durchführung handwerklicher Tätigkeitenin Virtual Reality(SmartFi)," and in part by GMH foundation "Stiftung Stahlwerk Georgsmarienhütte" (MW). This article was recommended by Associate Editor G. Serra. *(Corresponding author: Farbod Nosrat Nezami.)*

Shadi Derakhshan, Farbod Nosrat Nezami, Maximilian Alexander Wächter, Ashima Keshava, Hristofor Lukanov, Marc Vidal De Palol, and Gordon Pipa are with the Institute of Cognitive Science, University of Osnabrück, 49090 Osnabrück, Germany (e-mail: sderakhshan@uos.de; fnosratnezam@uos.de; mwaechter@uos.de; akeshava@uos.de; hlukanov@uos.de; mvidaldepalo@uos.de; gpipa@uos.de).

Artur Czeszumski is with the Institute of Cognitive Science, University of Osnabrück, 49090 Osnabrück, Germany, and also with the Department of Clinical Psychology, Vrije Universiteit Amsterdam, 1081 Amsterdam, The Netherlands (e-mail: aczeszumski@uos.de).

Peter König is with the Institute of Cognitive Science, University of Osnabrück, 49090 Osnabrück, Germany, and also with the Department of Neurophysiology and Pathophysiology, Center of Experimental Medicine, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany (e-mail: pkoenig@uos.de).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the ethics committee of the University of Osnabrück, and performed in line with the WMA Declaration of Helsinki.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2022.3168437.

Digital Object Identifier 10.1109/THMS.2022.3168437

I. Introduction

Autonomous vehicles (AVs) are the primary goal of most car manufacturers [1]. The development appears cumulative since more and more features of the driving task are automated in new cars [2]. One primary reason why AVs are of value is the possibility of eliminating human driving errors, which account for 93% of road accidents [3], [4]. Furthermore, AVs are safer because they are faster and more accurate in driving tasks as well as in the detection of objects and events [5]–[8]. Given the rapid and continuous improvement in technical developments in the field, there is no doubt that autonomous vehicles will have a significant impact on our society [9]. This may range from drastic reductions in greenhouse gas emissions to a decimation of traffic-related injuries. The introduction of AVs could lead to a possible reshaping of our existing cities' infrastructure [9]–[12]. Thus, introducing AVs into our daily lives seems to be a highly desirable objective.

Trust and acceptance of potential customers define the extent to which AVs are used for individual transportation [13], [14]. Current research indicates that there is a limited willingness among potential customers to use AVs [11], [15]–[18]. A variety of surveys have shown that most prospective buyers are unwilling to use an AV or to make full use of its functionalities [12], [19]. Primary reasons for the lack of trust and acceptance are the fear of system malfunctions and the hesitation to give full control to the car [12], [17], [20]. Hesitation among potential customers may result from low technology self-efficacy [21], which means people do not feel confident enough to operate an AV system [22]. Since trust and acceptance are shaped by knowledge and experience, the cause of such reluctance may be rooted in the lack of transparency, i.e., the AV decision-making foundation is not well defined. Lack of awareness of the cars' perceptions and the reasons behind the artificial agents' decisions has a direct impact on customers safety concerns [23],

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

[24]. It is, therefore, crucial to find measures to increase trust and acceptance of AVs.

According to the Technology Acceptance Model (TAM), perceived usefulness and ease of use are cognitive responses to new technologies and predict the intention of using them [25]. Consequently, a low intent to use this technology makes the future application of AVs questionable [13], [19], [26]. Belanche and colleagues have developed a research model [27] which expands the TAM by adding trust as a component. They found a causal relationship between trust and all three elements of the original TAM [27]. Therefore, trust can be regarded as a critical factor in the acceptance of a new technology [28]–[30]. Earlier studies investigated trust modulating factors, such as human–machine communication style, feedback, and anthropomorphic features in automation [31]–[34]. Hoff and Bashir [31] suggested that trust in an AV is an accumulation of personal tendencies, environment, and user's perception of the autonomous system. Lee and See [35] argue that the perceived homogeneity of communication style, feedback, and anthropomorphic features shape trust levels. The common statement in all these findings is that the user need to perceive the system as reliable and trustworthy. Moreover, previous research showed that excessive information about a car's operation is perceived as distracting or unpleasant [11]–[14], [19], [36]. The desired amount of information from an AV may be the key to understand the development of trust and, therefore, the acceptance of automated vehicles [37].

Research on trust as a high-level cognitive phenomenon is extensively studied with self-reported data. A review of Raats and colleagues of 258 trust-related experiments in AVs revealed that 84% used questionnaires as a method of assessment. Among them, only 4.7% of the studies used observations as a data-gathering tool [38]. Self-assessed data are often biased by self-perception or socially desired behavior, thus objective data can provide a better insight into the fields, such as trust [39]. To measure trust and acceptance objectively, we propose head movement analysis, since humans represent their cognitive states implicitly based on body language, facial expression, gaze direction, and movement of the head [40], [41]. Even though previous research has established the link between gaze shift and cognition [42], several studies showed that head rotation corresponds to the visual gaze [42], [43] and both coordinate cognitive processes [44], [45]. This coordination exists in the orientation, which means that the head and eyes move in the same direction. Thereupon, the head orientation provides information about the visual attention [43]. Behavioral data, such as head movements are unconscious, fine-grained, and continuous information that can be used to access cognitive processes like trust [46], [47].

To gain a better understanding of AV acceptance modulators and their representation in users head movements, we used a large-scale virtual reality (VR) [48]. We expected to find significant differences in attitudes among participants and significant differences across different age groups and genders. Additionally, we predicted differences in head movements between experimental conditions. We expected to find significant variance in head movement patterns and head angular velocity as an effect of transparent communication of an AV. We assumed that the self-reported acceptance in conjunction with head movements provides a more objective insight into the acceptance modulators.

II. Materials and Methods

We collected data from visitors in the German Ministry of Education over six months, and in a traveling exhibition (MS-Wissenschaft) during an entire summer. Participants experienced a 90-s drive in a virtual environment called Westdrive, which spans 2,5 km^2 and includes more than 100 cars and 150 pedestrians. Participants experienced a single trial in one of the three driving conditions. In the first condition, a fully autonomous car with an anthropomorphic voice assistant system (AVAS) provided information about critical traffic events and the corresponding car decisions. The second condition was an AV with a radio broadcast playing throughout the trial. In the third condition, a female TaxiDriver drove the participant through the city. Here, the TaxiDriver responded verbally to the surrounding traffic. We gathered objective and subjective data in the form of head orientation and head angular velocity, as well as by an adaptation of the TAM questionnaire [25].

Over the course of each trial, participants encountered three critical traffic events without being able to intervene (Fig. 1). The duration of the events was the time between the entry and the exit of the event objects from the participants' view. In the first event, a jogger crossed the road directly in front of the car. In the second event, a high-speed car took precedence at an intersection, and the third event included a pedestrian who was slowly crossing the street. The onset and end of the events were the same for all the participants in all experimental conditions. At neither event, the participants' cars hit any event objects. In the AVAS condition, the AV provided brief information on the critical event situation. The warning happened directly at the appearance of the critical traffic objects. The design of the events was based on previous research which indicated that feedback should include the reason an AV decides in a specific way [36]. Furthermore, additional information should be provided when interacting with vulnerable road users when their intentions are not clear but may influence the car's behavior [33]. These events were implemented to test the participants reactions as a passive passenger in critical situations. They were designed to test whether communication within the vehicle can affect behavioral responses and acceptance.

Before beginning the trial and data recording, participants were asked to make their own adjustment to the HMD. This adjustment phase was not limited in time nor was it taken into account in the experiment. Participants were subsequently made aware of the study procedure and the purpose of the study. Due to potential cybersickness symptoms, the participants were informed in the introduction that they could remove the HMD at any given time. In such cases, the data was excluded from the analysis.

The posttrial questionnaire includes three questions from the original TAM on perceived usefulness, ease of use, the intention of use, and one additional question on perceived trust. It also

Fig. 1. Three scripted critical events occurred during the ride from top to bottom: Pedestrians running on the street from left to right, fast cars cutting in the self-driving car path, and pedestrians walking in the middle of the road.

included questions on age, gender, aviophobia, driving experience, gaming hours per week, and the number of exposures to VR prior to the experiment. The questionnaire was answered on a Likert scale, with numbers from 0 (strongly disagree/dislike) to 100 (strongly agree/like) indicated by thumb icons of like and dislike.

The experimental setup consists of two HTC Vive pro HMDs and lighthouses version 1 to track head position and rotation, while participants were sitting in the car. The VR computers were equipped with Nvidia Geforce RTX 2080Ti GPUs, 16 Gb of RAM, and Intel Xeon W-2133 CPU @ 3.60Ghz core, resulting in an average frame rate of 25,2 fps. Additionally, the setup used two raspberry pies and touch monitors for web-based questionnaires. For analyses, Python 3.6, pandas 0.24.2, NumPy 1.16.4, Scipy 1.7.2, statsmodels 0.10.0, as well as SPSS 29 were used. All plots were created using Matplotlib 3.1.0 in combination with seaborn 0.9.0. Data-driven prepossessing on questionnaire data was performed with the OPTBIN algorithm [49] using histogram-based age binning.

A. *Analysis of the Data*

Head movement data were obtained from 26 750 participants and the questionnaire was answered by a fraction of them. Elimination of incomplete answers resulted in 7850 datasets.

First, we focused on analyzing the questionnaire data. Out of the full dataset, 4464 participants identified themselves as male and 3386 as female. Using optimal binning method [49], we divided participants into five age groups. The cleaned dataset consisted of 2812, 1513, 1883, 582, and 86 in the age groups <20 years, 21–40 years, 41–60 years, 61–80 years, and above 80 years, respectively. Under AVAS, TaxiDriver and RadioTalk conditions we recorded 2691, 2636, and 2509 datasets, respectively. The large number of participants in each bin permitted the use of regression-like inferential tests (i.e., MANOVA) due to their robustness against nonnormalities in large datasets [50].

To examine the effect of gender, age, and driving condition on the four aspects of the questionnaire, a one-way multivariate analysis of variance (MANOVA) was conducted. MANOVA tests the optimal linear combination of dependent variables for significant effects. We performed a one-way MANOVA for all four TAM aspects modeled based on gender, age, and driving condition. Pillai's trace test statistic uses the estimated F-Values to test significance, which is robust against nonnormalities. Therefore, Pillai's trace adds an extra layer of protection against false positives [51] and is a good choice to interpret the results. To understand how the different categories within each factor, e.g., male or female in gender, affect the four TAM aspects, we calculated a separate one-way analysis of the variance (ANOVA). Following, we calculated the different effect sizes (Cohen's D and Hedge's G) for each factor calculated using the estimated means and standard deviations reported for the category within this factor. Although both of these effect sizes are based on Cohen's suggestions, Hedge's G considers the sample sizes of the compared groups. Consequently, both effect sizes were used to interpret the findings. Further, the four TAM aspects of each participants were combined into one single value. Together with the MANOVA, we were able to make statements on how gender, age, and the condition affect the questionnaire scores.

However, ANOVA can only be calculated on a single independent variable. The best way to combine the four TAM aspects into one value is by multiplying each aspect's score for a given participant by a corresponding weight and adding them all together to get a single value. This acceptance score was calculated by performing a linear discriminant function for each factor that will yielded in a different raw coefficient for each TAM aspect concerning the given factor. The linear discriminant analysis (LDA) intends to find a linear combination of features that characterizes or separates two or more classes. It expresses the dependent variable as a linear combination of the independent variables that maximizes the group differences within the dependent variable [52]. The raw discriminant function coefficients can be used as weights to calculate the four TAM aspects into one independent number, which we can call acceptance score.

Then we proceed to the analysis of objective behavior. A head-mounted HMD measured the orientation and position of the participant's head in the virtual environment. We define the orientation of the head in a frame of reference attached to the car. Since most interesting visual detail was placed near the ground level and all the dynamic objects of the virtual city moved along the horizontal axis, we focused on the orientation along the horizontal plane. Moreover, we compared the head angular velocity, meaning the change in head orientation degree over time. To examine the differences in conditions, we used one-way ANOVA followed by the *post hoc* Tukey honest significant difference (HSD) test. The Tukey HSD compares pairs of means to detect which of the group's means differs from the others (Meandiff). With this test, we could define the separate condition that causes differences in orientation and angular velocity in a given time point [53]. Additionally, we calculated the Pearson correlation between head angular velocity and TAM scores for each questionnaire item to verify the consistency of subjective and objective measures.

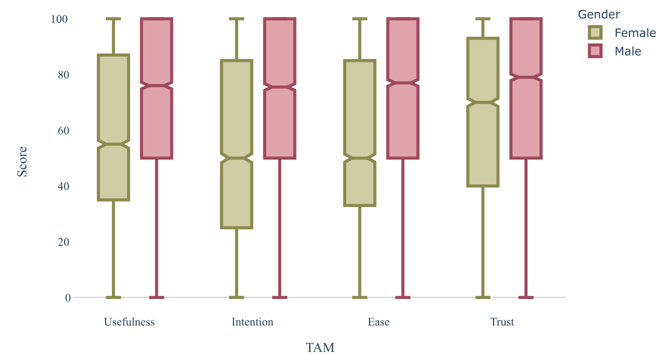
III. Results

A. Questionnaire Results

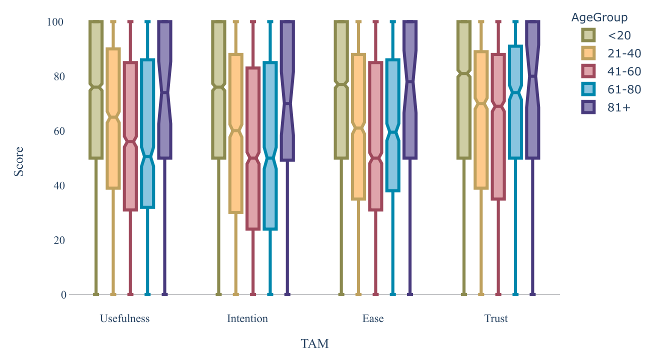
The data of the simplified TAM questionnaire from 7850 participants showed a positive correlation of $r>0.4$ between the questionnaire items. Hence, these items have to be analyzed together in form of multivariate dependent variables. In order to validate the assumptions, a Levene's test was performed. If the test was significant we would assume a violation of variance homogeneity in the groups. Levenes's test resulted in F-values of 1.369 for perceived Usefulness ($p = 0.089$), 2.333 for Ease of use ($p < 0.001$), 1.459 for Intention of use ($p = 0.053$), and 1.443 for Trust ($p = 0.058$). Considering the large sample size, known to reduce p-values in Levene's test, a further check of the covariance matrices for the dependent variables of the TAM concerning the main factors of gender, age group, and condition has been done. We found homogeneity of covariances, as assessed by Box's test (p >. 001). Together, Levene's test and the covariance matrices provide essential evidence for the validity of the assumptions of MANOVA. Out of the four different null hypothesis tests of the multivariate analysis, Pillai's trace was chosen due to its known robustness toward nonnormalities in the data [54]. Therefore, the multivariate analysis of variance is the prime analysis method [55], [56].

To gain deeper insights into how gender, age, and condition affect the TAM factors, a LDA was used to extract each independent variable's weighted influence. LDA tries to find a set of coefficients that will maximize the separability within the given independent variable. These coefficients were used to interpret the influence of each independent variable on each of the modulator factors of the TAM.

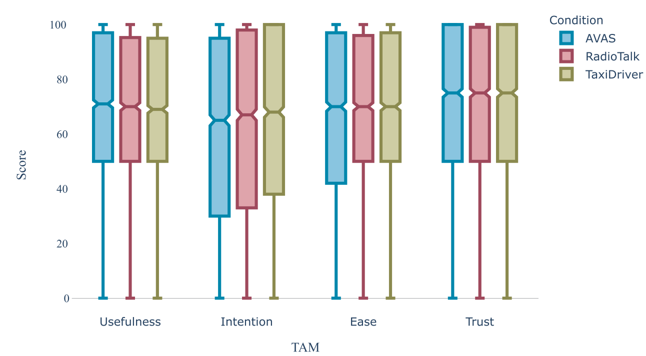
1) Effect of Gender: First analysis checked for differences between male and female participants regarding the acceptance scores. In order to find out the influence of gender on acceptance, we performed an MANOVA with a follow-up LDA for gender. The Pillai's trace resulted in 0.00293 (F(4,7835) = 4.761, $p <$ 0.001) showing that there is a significant effect of gender on overall acceptance. The follow-up LDA showed that females have a lower score based on the observed discriminant coefficients. The resulting coefficients were -0.33 for the intention of use, -0.06 for perceived usefulness, -0.60 for perceived ease of use, and -0.18 for trust [Fig. 2(a)], all with a medium effect size (Cohen's D $= 0.45$). Additionally, the LDA showed that perceived usefulness and trust were less affected by gender than the intention of use and the perceived ease of use (Fig. 2 a).



(a)



(b)



(c)

Fig. 2. Descriptive categorical plot of the mean questionnaire answers for each of the main factors (a) gender, (b) age group, and (c) condition.

These findings indicate that females and males have an almost equivalent attitude toward the perceived usefulness but differ in the perception of ease of use and, consequently, the intention of using self-driving cars. Thus, we interpreted that females anticipate difficulties in handling and, therefore, score lower in the intention to use.

2) Effect of Age Group: As a next step, we investigated what influence age had on the answers in the questionnaire. Similar to the effects of gender, the result of the MANOVA was paired with a follow-up LDA for the age group to find the influence of the TAM items. The resulting Pillai's trace of 0.04561 (F(16,313) = 24.107, $p < 0.001$) indicated a significant effect of age group on the overall acceptance. LDA resulted in discriminant coefficients of -0.27 for the intention of use, -0.19 for perceived usefulness,

-0.46 for perceived ease of use, and -0.30 for trust. These results showed that age has a negative effect on all TAM items. The age group under 20 years showed high scores in all questions, with medium effect sizes compared to the age group between 20–40 (Cohen's D = 0.50) and 61–80 (Cohen's D = 0.43) (for the full list of the effect sizes, see Appendix A–Table I). Similar to the analysis of gender, perceived ease of use was affected by age the most [Fig. 2(b)]. Together with the decreased intention of use, it can be stated that older adults anticipate hardships in using this technology, therefore, showing lower scores in the intention to use. However, the overall acceptance scores increased again beyond 80 years of age, especially in the perceived usefulness [Fig. 2(b)]. This is also reflected in smaller effect sizes between the age group below 20 and above 80 years (Cohen's D = 0.18) (SI Table1). Concluding, data showed the highest acceptance in the age group below 20 years, with a general decrease of acceptance until 80 years.

3) Effect of Condition: A central hypothesis of the study was, that compared to a traditional taxi the acceptance level of AVs is reduced, but can be partially recovered by making the decisions of the AV transparent. The result of MANOVA showed that the condition had a significant effect with Pillai's trace of 0.00259 (F(8,15672) = 2.541, $p = 0.009$). The LDA for condition resulted in coefficients of -1.12 for the intention of use, 0.99 for perceived usefulness, -0.33 perceived ease of use, and 0.44 for trust, with overall small effect sizes in all comparisons (Cohen's D = <0.11) (Appendix A-Table II). While the effect on trust and ease of use is negligible between conditions, differences could be found in intention to use and perceived usefulness. The AVAS condition had a slightly higher median score in the perceived usefulness of 71 compared to 69 in the TaxiDriver [Fig. 2(c)]. Here, the AVAS condition resulted in a lower median score of 65 than the TaxiDriver with a median of 68 [Fig. 2(a)]. We concluded that there were no averse effects of the condition on the ease of use like in gender and age. Still, there was a negative effect on the intention to use such technology independently of gender and age. These results already accommodated that additional factors besides age and gender negatively influenced the intention of use. This observed effect could also be due to technology aversion, which had already been mentioned in the effects of age and gender. It can be summarized that there was a small positive effect of in-car communication methods on accepting AVs regarding the ease of use and a small negative effect regarding the intention of use.

4) Interaction Effect of Gender and Age Group: In examining the effects of gender, age, and condition, it became clear that these factors separately did not explain all variance observed in the data. There was a significant interaction effect of gender and age group with Pillai's trace of 0.00498 (F(16,31352) = 2.441, $p = 0.001$). According to the follow-up LDA, there was a negative effect for the intention of use and perceived ease of use (both -0.73) and a positive effect on the perceived usefulness (0.22) and trust (0.55) in the questionnaire items. Here, the effect sizes were most notably between the age groups 21–60 years compared to under 20 years and above 60 for each gender (Appendix A-Table III). These results support findings of the previous analyses on gender and age. In addition, it could be

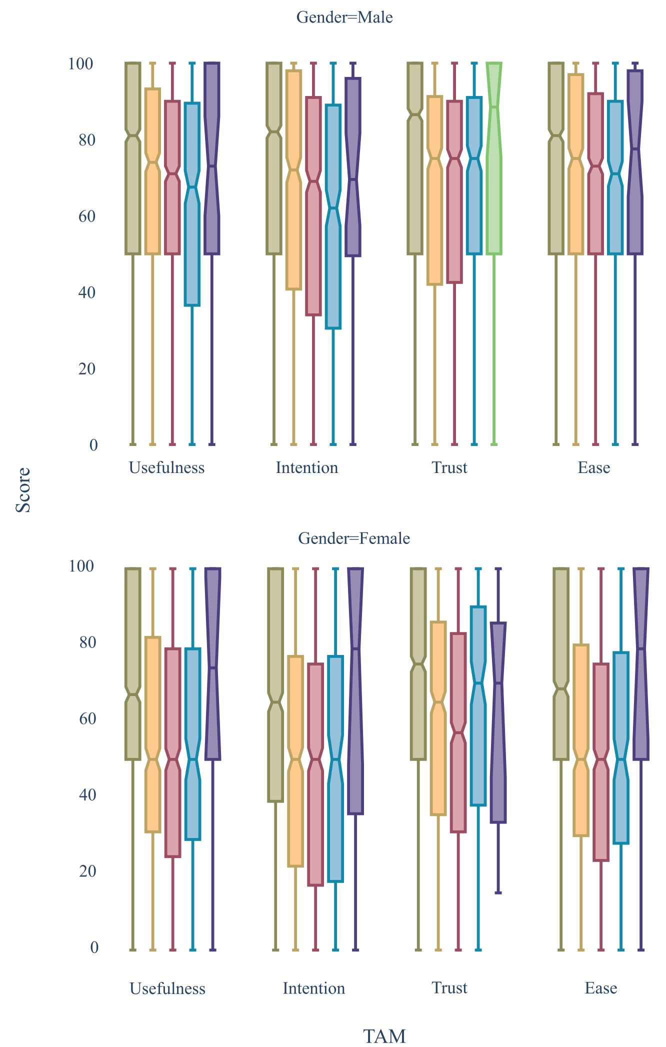
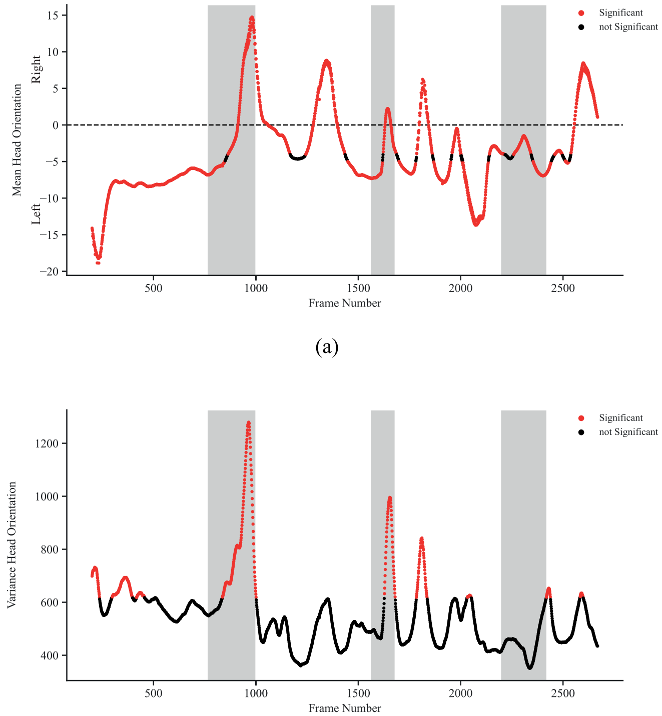


Fig. 3. Mean of answers for questions regarding the usefulness, intention, trust, and ease for age group and gender combined.

shown that the interaction of gender and age had a significant influence on the acceptance of AVs. In addition, the largest effect sizes (0.5 < Cohens' D <= 0.9) resulted by comparing female participants in the age between 21–60 years against the male participants in the same range. Participants below 20 years had the highest TAM scores, and females between the ages of 21–60 years showed the lowest TAM scores. Although there is was a decrease in all TAM factors in both genders for ages between 21–60 years old compared to the below 20 years, female participants showed stronger decreases in TAM scores (Fig. 3). This accounts especially for the intention of use and perceived ease of use. Once again, as age increases for people between ages 21–80 years, we can also observe TAM scores. In conclusion, gender and age group interaction significantly affect all TAM factors, specifically negative influences on the intention of use and perceived ease of use for AVs, but a positive effect on perceived usefulness and trust. This means that although the AV was seen as useful and trustworthy, there were still other hidden factors that decreased the ease of use and the intention to use it. Consequently, the demographic factors of age, gender, and the



(a)

(b)

Fig. 4. Time intervals of significant differences in head orientation after the permutation test ($n = 1000$). Shaded areas represent the critical traffic event intervals. (a) Mean of head orientation over all subjects. Each point indicates an average of the head orientation across the participant within each frame. (b) Variance of head orientation over all subjects. The red areas indicate the intervals where there was a difference in head orientation between all participants. Please note that the data are collapsed over condition.

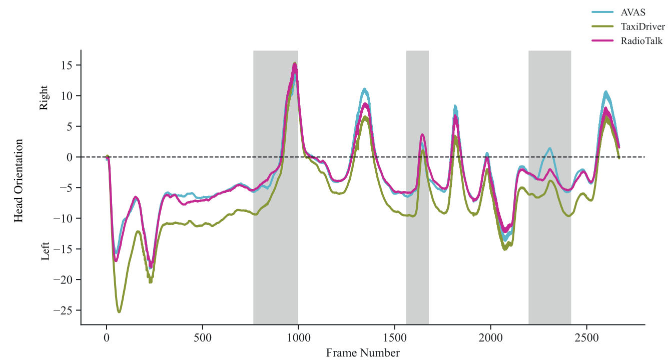


Fig. 5. Mean head orientation in each frame divided in three conditions. The positive and negative values of the mean orientation relate, respectively, to the right and left directions. Shaded areas represent the critical traffic event intervals.

interaction of these two, have much more impact on the items of the TAM questionnaire. The positive effects of a self-explanatory AV were not sufficient to compensate for the negative influence of demographics on ease of use and, accordingly, intention to use.

B. Behavioral Results

1) Identification of Critical Events: As a first step in the analysis of head movement, we tried to determine if the behavior of the participants differs during the critical events from the baseline parts of the drive. The initial analysis was conducted independently of the driving condition. We considered the mean and variance of head orientation over all participants as the relevant dependent variables. Collapsing the data over conditions, we tested whether the mean of head orientation in each frame was significantly different from the distribution resulting from a permutation over time (permutation test). The head orientation differed significantly from the baseline at the end on the first and second and at the very end of the third event. (Fig. 4). Further, we observed differences early in the trial, when participants were intensely looking around inside the car. There were also three other significant intervals observed over the trial. During these intervals, pedestrians were visible on the sidewalk in crowded areas of the city. We assume that this is related to a need for information to assess the situation. A last period of deviant head orientation is observed at the very end of the drive, when participants have prepared to get out of the car in a congested area. By applying this method we are confident that an additional measure of head movements is a valid approach to enhance subjective data. Overall, compared to the baseline head orientation, the three critical events revealed significant differences in the participant behavior regardless of the condition effect. These differences were not limited to the critical event intervals but identified in additional areas of the trial.

2) Effect of Condition: In the next stage, we looked at the extent to which the observed variance was associated with the effect of the condition. We investigated whether participants head movements objectively varied between conditions. Differences in head orientation were seen as indicators of participants' reaction to the environment in different experimental conditions. In visualizing the head movement data, we observed differences in the mean head orientation, over large portions of the drive and during the critical traffic events. (Fig. 5). In order to determine if these differences were significant, a one-way ANOVA based on head orientation was calculated for each frame as a dependent variable and applied a *post hoc* comparison of Tukey HSD in those significant intervals. The ANOVA result showed significant differences in head orientation between the three conditions for most of the frames ($F > 10, p < 0.05$) (Fig. 6). Specifically, the TaxiDriver condition was significantly different from the other two on much of the drive. The *post hoc* comparison revealed larger mean difference (Meandiff) for TaxiDriver compared to AVAS (i.e., Meandiff $= 2.07$, $p = 0.001$ for frame $= 1300$) and RadioTalk (i.e., Meandiff $= 1.77$, $p = 0.001$ for frame $=$ 1300) conditions. This outcome was most often constant during the experimental trial, including all three critical events. At the beginning of the first and the second events, no significant differences were observed between the RadioTalk and AVAS conditions were found. In the third event we noted differences across all three conditions. At this point, the participants got the lowest degrees of the head orientation in the AVAS condition, and the highest degree in the TaxiDriver condition. In particular, we observed a higher mean head orientation in the TaxiDriver condition at the beginning of the critical traffic events. This means that the distribution of head orientations in the angular

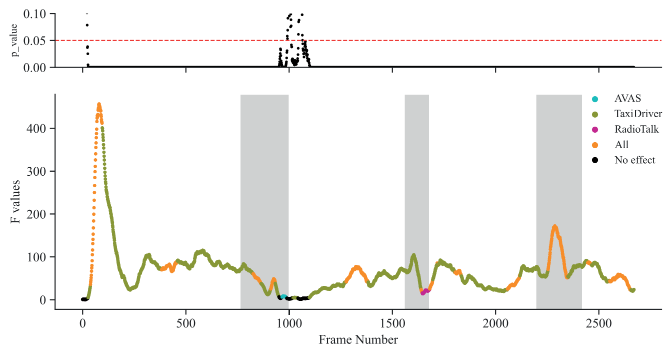


Fig. 6. Time intervals of significantly different behavior between the three conditions. The graph depicts the P and F values of one-way ANOVA overall the experimental trial. Each dot shows the original F value of each frame. The red dash indicates the significance threshold ($p < 0.05$). Shaded areas represent the critical traffic event intervals. The result of Tukey's post hoc comparison is represented by different colors. Each color shows the significant variable mean(s) in cross-check.

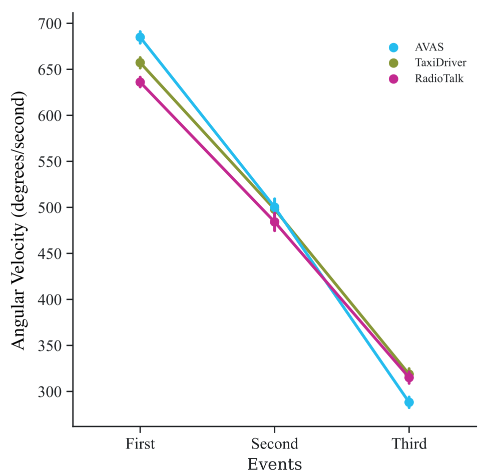


Fig. 7. Head angular velocity for the critical event intervals. The y-axis refers the rotation change divided by the number of event frames.

space were wider in the TaxiDriver condition compared to the two autonomous conditions (Fig. 5).

3) Head Angular Velocity: To gain deeper insights into the participants' head movement behavior, we calculated the magnitude of change in head orientation over time as the angular velocity. We quantify the absolute value of the angular velocity in frame n for each critical traffic event separately ($\omega_n = |\theta_n - \theta_{n-1}|/\Delta t$) where θ_n is the head orientation in frame n and $\Delta t = 0.04\tilde{}s$ based on the experiment's overall average frame rate. Angular velocity analysis showed that in the AVAS condition, participants rotated their heads significantly faster only during the first critical traffic event [$F(2,24447) = 71.35, p < 0.01$]. In the second critical traffic event, no significant differences were observed between the conditions [$F(2,24447) = 2.8$, $p = 0.06$]. In the third critical traffic event, the angular velocity in AVAS was significantly lower than in the other two conditions [$F(2,24447) = 29.06$, $p < 0.01$] (Fig. 7). Overall, the data revealed that the angular velocity of head movement decreased during the experimental trial in all three conditions. However, the AVAS condition reduced the head's angular velocity to a greater degree than the other autonomous condition. With angular velocity analysis, we were able to show that participants' behavior changed as an effect of self-explaining AV over time.

C. Questionnaire Comparison

The head angular velocity was an illustration of the participant's head movements behavior during the trial. The calculation of the relationship between angular velocity and TAM items allowed us to determine if the self-assessment of participants was expressed in their prior behavior during the experimental trial. We used Pearsons' correlation for the participant's absolute head angular velocity over the entire trial and the participants' respective TAM item scores. The analysis showed a positive correlation between the head angular velocity and all TAM item scores for all three conditions (Fig. 8). The correlation between the angular velocity and trust was lower than its correlation to other TAM items. Along with the previous finding in the analysis of the questionnaire, the mismatch between trust and the other items of the questionnaire was demonstrated in the correlation between the items and the angular velocity. The dissociation between trust and the other questionnaire items suggests that trust is not an ideal item in self assessments via a questionnaire. This allegation is supported by the mismatch in the self-assessment together with the objective behavioral data. As a result, we argue that the objective behavioral data was able to reflect the findings of the TAM questionnaire.

IV. Discussion

The present study revealed that self-reported acceptance in conjunction with objective observation, provide a better understanding of modulating acceptance factors. The results indicated that, subjective data from a postexperimental questionnaire and objective data from head movements during the experimental trial were largely congruent. The outcome of the study on gender, age, and the effect of condition on the overall acceptance showed less acceptance of female participants toward AV than among males. However, this effect is even more pronounced in the intention to use AVs. The results also suggested that people under the age of 20 have the highest acceptance toward AV, declining gradually with age, while increasing again above the age of 80. Concerning the effect of a self-explaining AV, we found a small positive effect in the ease of use and a small negative effect regarding the intention of use. However, the age, gender, and the interaction of these two factors have a substantially greater impact on the questionnaire results. Therefore, the positive effects of a self-explanatory AV are not sufficient to compensate for the negative influence of demographic data on ease of use and the intention to use. It could be demonstrated that participants' head orientation differed between the conditions by analyzing the data on head movements. In particular in the TaxiDriver condition, we observed significant differences across the entire drive with additional differentiation between conditions during the critical events. Further, we observed a decline in the head's

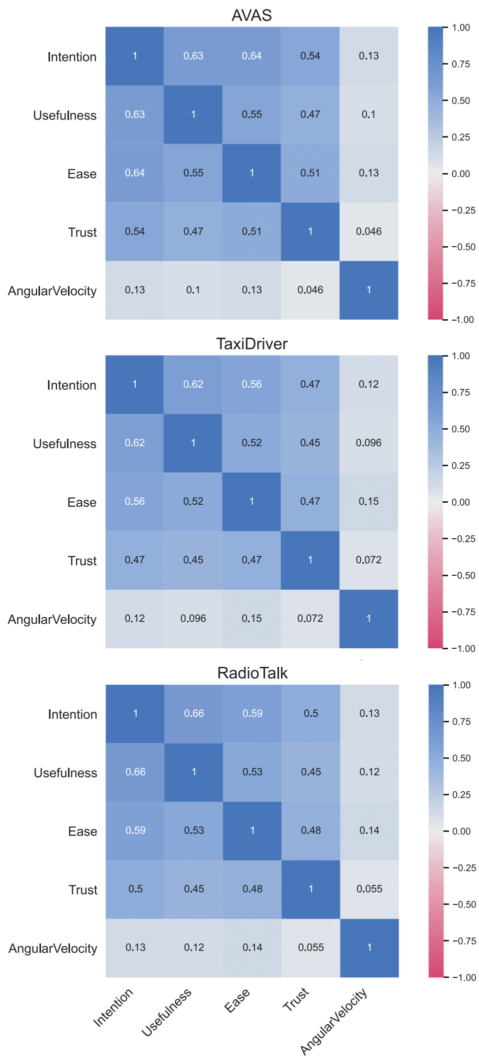


Fig. 8. Summary of the Pearson correlation between the head angular velocity and the TAM questionnaire items. The correlation p value for Intention, usefulness, ease of use and trust are as follows for each condition. (a) AVAS (Intention: $p<0.001$, Usefulness: $p<0.0010$, Ease of use: $p<0.001$, Trust: $p<0.01$). (b) RadioTalk (Intention: $p<0.001$, Usefulness: $p<0.001$, Ease of use: $p<0.001$, Trust: $p<0.001$). (c) TaxiDriver (Intention: $p<0.001$, Usefulness: $p<0.001$, Ease of use: $p<0.001$, Trust: $p<0.001$).

angular velocity over time for all conditions. The latter effect was most substantial in the AVAS condition. Finally, correlate the magnitude of the participant's head angular velocity to the TAM scores showed a significant relationship between acceptance as a combination of the TAM factors, which was weaker for the trust factor.

Earlier studies were primarily based on the responses of potential users identified through questionnaires [13], [33], [38]. However, behavioral data are not as susceptible as questionnaire answers and can be used to validate possible self-assessments [25]. The presented study could show a dissociation of the self-assessed trust from other TAM items, in particular with the intention to use. This observation contrasts with previous research, such as that of Belanche [27]. We demonstrated that self-assessments are strongly modulated by the demographic factors, such as age and gender, as well as the interaction of these two factors. Behavioral data confirmed the dissociation of trust and intention, by showing a connection between head movements and scores in intention, ease of use, and perceived usefulness items. Therefore, we argue that inclusion of behavioral data are a valid approach to better understand underlying factors of acceptance and justify potentially flawed subjective data. This is due to the fact that head movements can be considered as part of nonverbal communication among humans [57] that contains information about the participant's emotions and intentions [58]. For instance, the angular velocity of the head and its acceleration were higher during negative affects [59]. The combination of subjective and objective data sources allows data from the questionnaire to be validated. In conclusion, behavioral data could be described as an important resource that can be used to validate investigations into the technology acceptance model and its underlying factors.

Due to the nature of the experiment in a public exhibition and a large number of visitors, we used a simplified version of the technology acceptance questionnaire. Thus, it may fail to capture the full aspect and spectrum of factors that modulate acceptance, such as the technology self-efficacy, which might play a critical role in perceived ease of use. Moreover, the questionnaire was translated into German, and we were not able to validate it prior to using it in the experiment. As a result, part of the variance in the data could be caused by the translation. However, such an effect is considered to be minuscule and negligible since our main findings aligned with those of the previous work [12], [23], [60], [61]. Certainly, when it comes to trust in AVs, there are more modern and validated methods, such as STS-AD [62] and Jian scale [63], which could better explain the underlying factors of trust. Therefore, adapting such methods will undoubtedly benefit future studies to explore the trust in AVs. However here our initial goal is to investigate acceptance in AVs and to design behavioral methods to study acceptance and trust. Given that the Jian and STS-AD scales are questionnaire-based measures, suffers the same issues arising from self-reporting nature of questionnaires. Due to the simplified nature of the study, we cannot directly address and analyze the underlying information processing that influences the attitude. Nonetheless, we are confident in making informed statements because of the magnitude of the effects of a vast dataset. Additionally, it is also possible that cybersickness has influenced TAM scores and head movement data. Nevertheless, we tried to control the motion and cybersickness as much as we could in this trial. To reduce the probability of emerging cybersickness, we have tried to pay attention to certain precautions in the creation of the environment. Among them falls a bigger static frame for the participants as the car interior. In addition, we only used a low-speed environment, with no tight corners, to minimize cybersickness [64]. Furthermore, we acknowledge that a more precise measurement instrument, such as eye trackers would have improved the analysis and the findings. However, once again, the nature of the experiment and the absence of on-site experimenters made it impossible to use such methods. Another criticism might be that the experimental time was limited to 90 s, and each participant observed a single experimental condition. However, this experiment already provides an opportunity to investigate participants' acceptance in communicative AVs. Additionally, the vast amount of data collected through the experiment allowed for entirely data-driven analyses both for questionnaire and behavioral data. Consequently, the results of this study are valuable in understanding public acceptance of AVs and the importance of objective measures.

Despite these limitations, we are confident that we were able to demonstrate the effect of a self-explaining AV based on subjective and objective data. As mentioned earlier, previous research has explained trust as a combination of the communication style, feedback, and the anthropomorphic characteristics of the AV [27], [36]. In contrast, Hoff and Bashir argued that trust is largely shaped by the users' personality traits [31]. This is supported by new findings in real driving scenarios, where personality traits were identified as relevant factors of trust [65], and were only out weighted by the actual driving performance. These factors are summarized under "dispositional trust," which comprises age and gender, and personality traits. Consistent with previous research our study could show that the demographic factors have a greater impact on acceptance compared to the characteristics of AVs.

Nevertheless, our findings are not generalizable across demographic groups, since their communication needs are different; while we see a positive influence of the talking car in one group, the second group may regard the in-vehicle information as excessive and distracting. Thereafter, the user-specific communication could increase trust in doubtful users, making them more confident to properly operate such a system since it might be able to increase the knowledge of the autonomous system. However, further investigation is required using more extensive questionnaires to examine other acceptance modulators, specifically trust in combination with more objective measuring instruments, such as eye tracking. In the end, we argue that user specific in-vehicle communication can be helpful in creating guidelines for the development of a safer and inclusive future of autonomous mobility.

Appendix A
Calculated Effect Sizes for Each Significant Factor

TABLE I
EFFECT SIZES BETWEEN DIFFERENT AGE GROUPS ON INTENTION TO USE, PERCEIVED USEFULNESS, AND PERCEIVED EASE OF USE AND TRUST

Age group	Below 20	20–40	40–60	60–80	Above 80
below 20	0	0.40	0.50	0.43	0.17 / 0.2
20 - 40	0.40	0	0.10	0.01	0.18 / 0.20
40 - 60	0.50	0.10	0	0.08	0.28
60 - 80	0.43	0.01	0.08	0	0.20 / 0.22
Above 80	0.17 / 0.2	0.18 / 0.20	0.28	0.20 / 0.22	0

Numbers in the table present the Cohen's D and in the case of difference hedges G.

TABLE II
EFFECT SIZES BETWEEN DIFFERENT CONDITION ON INTENTION TO USE, PERCEIVED USEFULNESS, AND PERCEIVED EASE OF USE AND TRUST

Condition	AVAS	RadioTalk	TaxiDriver
AVAS	0	0.05	0.11
RadioTalk	0.99	0	0.06
TaxiDriver	0.11	0.06	0

Numbers in the table present the Cohen's D and in the case of difference hedges G.

TABLE III
EFFECT SIZES BETWEEN DIFFERENT COMBINATION OF GENDER AND AGE GROUP ON INTENTION TO USE, PERCEIVED USEFULNESS, PERCEIVED EASE OF USE AND TRUST

gender = male					
Gender/Age Group	below 20	20 - 40	40 - 60	60 - 80	80+
Male / below 20	0	0.30	0.36	0.39	0.28 - 0.32
Male / 20 - 40	0.30	0	0.06	0.07	0
Male / 40 - 60	0.36	0.06	0	0.01	0.05
Male / 60 - 80	0.39	0.07	0.01	0	0.06
Male / 80+	0.28 - 0.32	0	0.05	0.06	0
Female / below 20	0.31	0.01	0.05	0.06	0.04
Female / 20 - 40	0.73	0.42	0.36	0.35	0.39 - 0.41
Female / 40 - 60	0.90	0.58	0.51	0.51	0.53 - 0.56
Female / 60 - 80	0.78 - 0.80	0.46	0.40	0.40	0.42 - 0.44
Female / 80+	0.29 - 0-33	0	0.03	0.05	0.01
gender = female					
Gender/Age Group	below 20	20 - 40	40 - 60	60 - 80	80+
Male / below 20	0.31	0.73	0.9	0.78 - 0.80	0.29 - 0-33
Male / 20 - 40	0.01	0.42	0.58	0.46	0
Male / 40 - 60	0.05	0.36	0.51	0.40	0.03
Male / 60 - 80	0.06	0.35	0.51	0.40	0.05
Male / 80+	0.04	0.39 - 0.41	0.53 - 0.56	0.42 - 0.44	0.01
Female / below 20	0	0.42	0.57	0.46	0.01
Female / 20 - 40	0.42	0	0.15	0.03	0.37 - 0.40
Female / 40 - 60	0.57	0.15	0	0.11	0.51 - 0.55
Female / 60 - 80	0.46	0.03	0.11	0	0.41 - 0.44
Female / 80+	0.01	0.37 - 0.40	0.51 - 0.55	0.41 - 0.44	0

Numbers in the Table Present the Cohen's D and in the Case of Difference Hedges G.

Acknowledgment

The authors would like to thank all involved members of the WestDrive team who helped building software and hardware as well as data acquisition. They would like to thank J. M. Pingel and S. Kim for helping to organize and commenting the experimental code, again J. M. Pingel for maintaining hardware during the duration of the exhibition. They would also like to thank D. Gün and T. Klein for implementing the supplementary material into the exhibitions, as well as P. Spaniol for designing and building the exhibitions and, lastly, M. V. de Palol and H. Lukanov for implementing the Questionnaire, supplementary material, and server structure. Without the support of all these wonderful members of the institute, this project would not have been realized.

A. *Authors Contribution*

FNN and MAW developed, designed and conducted the experiment as well as analysis of the questionnaire data. SD analyzed the head-tracking data and contributed to the writing process. SD, FNN, and MAW share first authorship. AK and AC assisted with data analysis as well as the review of the manuscript, HL and MV has developed the questionnaire used in the experiment. PK and GP supervised the project and reviewed the manuscript and share senior authorship.

B. *Conflict of Interests*

The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

C. Datasets

Code of the entire experiment conducted in this article is available at https://gitlab.com/fnnezami/project-westdrive Gitlab repository under a creative common license. Furthermore all analysis scripts and output resuts, as well as raw data and demo video are available at https://osf.io/udhg7/?view_only=88cd3f61e10b409d8064c13518105486 OSF under a creative common license.

References

[1] A. Hars, "Top misconceptions of autonomous cars and selfdriving vehicles. Thinking outside the box," *Innovation Briefs Issue 2016-09 (Version 1.3)*, 2016. [Online]. Available: https://www.inventivio.com/innovationbriefs/2016-09/Top-misconceptions-of-self-driving-cars.pdf

[2] Y. Dajsuren and M. van den Brand, "Automotive software engineering: Past, present, and future," in *Automotive Systems and Software Engineering: State of the Art and Future Trends*, Y. Dajsuren and M. vanden Brand, Eds., Berlin, Germany: Springer, 2019, pp. 3–8.

[3] T. Allahyari *et al.*, "Cognitive failures, driving errors and driving accidents," *Int. J. Occupational Saf. Ergonom.*, vol. 14, no. 2, pp. 149–158, 2008.

[4] K. Pandey, P. Fulzele, R. Singh, P. Kumar, and S. Singh, "Predicting and preventing fatal crashes," in *Proc. 12th Int. Conf. Contemporary Comput.*, Aug. 2019, pp. 1–6.

[5] SAE International, "J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems—SAE international," 2014. [Online]. Available: https://www.sae.org/standards/content/j3016_201401/

[6] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data," *Remote Sens.*, vol. 13, no. 1, 2021, Art. no. 89.

[7] A. Papadoulis, M. Quddus, and M. Imprialou, "Evaluating the safety impact of connected and autonomous vehicles on motorways," *Accident Anal. Prevention*, vol. 124, pp. 12–22, 2019.

[8] B. Schoettle, "Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles," Univ. Michigan, Ann Arbor, MI, USA, Rep. no. SWT-2017-12, 2017.

[9] A. Chehri and H. T. Mouftah, "Autonomous vehicles in the sustainable cities, the beginning of a green adventure," *Sustain. Cities Soc.*, vol. 51, 2019, Art. no. 101751.

[10] A. Z. Benleulmi and T. Blecker, "Investigating the factors influencing the acceptance of fully autonomous cars," in *Proc. Hamburg Int. Conf. Logistics*, 2017, pp. 99–115.

[11] M. Ryan, "The future of transportation: Ethical, legal, social and economic impacts of self-driving vehicles in the year 2025," *Sci. Eng. Ethics*, vol. 26, no. 3, pp. 1185–1208, Jun. 2020.

[12] K. Othman, "Public acceptance and perception of autonomous vehicles: A comprehensive review," *AI Ethics*, vol. 1, pp. 355–387, Feb. 2021.

[13] D. Howard and D. Dai, "Public perceptions of self-driving cars: The case of Berkeley, California," in *Proc. Transp. Res. Board 93rd Annu. Meeting*, 2014, pp. 1–16.

[14] R. Krueger, T. H. Rashidi, and J. M. Rose, "Preferences for shared autonomous vehicles," *Transp. Res. Part C, Emerg. Technol.*, vol. 69, pp. 343–355, Aug. 2016.

[15] M. Kyriakidis, R. Happee, and J. C. F. de Winter, "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents," *Transp. Res. Part F. Traffic Psychol. Behav.*, vol. 32, pp. 127–140, 2015.

[16] C. Lee, C. Ward, M. Raue, L. D'Ambrosio, and J. F. Coughlin, "Age differences in acceptance of self-driving cars: A survey of perceptions and attitudes," in *Human Aspects of IT for the Aged Population. Aging, Design and User Experience*, Berlin, Germany: Springer, 2017, pp. 3–13.

[17] C. Lee, B. Seppelt, B. Reimer, B. Mehler, and J. F. Coughlin, "Acceptance of vehicle automation: Effects of demographic traits, technology experience and media exposure," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2019, pp. 2066–2070.

[18] M. R. Ward, C. Lee, L. D'Ambrosio, J. F., and M. Coughlin, "Acceptance of automated driving across generations: The role of risk and benefit perception, knowledge, and trust," in *Human-Computer Interaction. User Interface Design, Development and Multimodality*, vol. 10271, Berlin, Germany: Springer, 2017, pp. 254–266.

[19] A. Rezaei and B. Caulfield, "Examining public acceptance of autonomous mobility," *Travel Behav. Soc.*, vol. 21, pp. 235–246, Oct. 2020.

[20] P. Szikora and N. Madarász, "Self-driving cars—The human side," in *Proc. IEEE 14th Int. Sci. Conf. Inform.*, 2017, pp. 383–387.

[21] S. J. Czaja, J. Sharit, N. Charness, A. D. Fisk, and W. Rogers, "The center for research and education on aging and technology enhancement (CREATE): A program to enhance technology for older adults," *Gerontechnology*, vol. 1, no. 1, pp. 50–59, Sep. 2001.

[22] M. König and L. Neumayr, "Users' resistance towards radical innovations: The case of the self-driving car," *Transp. Res. Part F. Traffic Psychol. Behav.*, vol. 44, pp. 42–52, Jan. 2017.

[23] J. Koo, D. Shin, M. Steinert, and L. Leifer, "Understanding driver responses to voice alerts of autonomous car operations," *Int. J. Veh. Des.*, vol. 70, no. 4, pp. 377–392, 2016, Art. no. 377392.

[24] Y. Forster, F. Naujoks, and A. Neukum, "Increasing anthropomorphism and trust in automated driving functions by adding speech output," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 365–372.

[25] F. D. Davis and V. Venkatesh, "A critical assessment of potential measurement biases in the technology acceptance model: Three experiments," *Int. J. Human-Comput. Stud.*, vol. 45, no. 1, pp. 19–45, Jul. 1996.

[26] L. T. Bergmann *et al.*, "Autonomous vehicles require socio-political acceptance-an empirical and philosophical perspective on the problem of moral decision making," *Front. Behav. Neurosci.*, vol. 12, 2018, Art. no. 31.

[27] D. Belanche, L. V. Casaló, and C. Flavián, "Integrating trust and personal values into the technology acceptance model: The case of e-government services adoption," *Cuadernos de Economía y Dirección de la Empresa*, vol. 15, no. 4, pp. 192–204, Oct. 2012.

[28] C. Lee and J. F. Coughlin, "PERSPECTIVE: Older adults' adoption of technology: An integrated approach to identifying determinants and barriers," *J. Prod. Innov. Manage.*, vol. 32, no. 5, pp. 747–759, Sep. 2015.

[29] P. Wintersberger and A. Riener, "Trust in technology as a safety aspect in highly automated driving," *i-com*, vol. 15, no. 3, pp. 297–310, Dec. 2016. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/icom-2016-0034/html

[30] M. Lüders and P. B. Brandtzæg, "'my children tell me it's so simple': A mixed-methods approach to understand older non-users' perceptions of social networking sites," *New Media Soc.*, vol. 19, no. 2, pp. 181–198, 2017.

[31] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human Factors, J. Hum. Factors Ergonom. Soc.*, vol. 57, no. 3, pp. 407–434, May 2015.

[32] P. Wintersberger, A.-K. Frison, A. Riener, and T. von Sawitzky, "Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality," *Presence Virtual Augmented Reality*, vol. 27, no. 1, pp. 46–62, Mar. 2019. [Online]. Available: https://direct.mit.edu/pvar/article/27/1/46-62/96082

[33] P. Wintersberger, H. Nicklas, T. Martlbauer, S. Hammer, and A. Riener, "Explainable automation: Personalized and adaptive UIs to foster trust and understanding of driving automation systems," in *Proc. 12th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2020, pp. 252–261.

[34] B. D. Seppelt and J. D. Lee, "Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation," *Int. J. Human-Comput. Stud.*, vol. 125, pp. 66–80, 2019.

[35] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[36] J. Koo, J. Kwac, W. Ju, M. Steinert, L. Leifer, and C. Nass, "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance," *Int. J. Interact. Des. Manuf.*, vol. 9, no. 4, pp. 269–275, Nov. 2015.

[37] N. Du *et al.*, "Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload," *Transp. Res. Part C, Emerg. Technol.*, vol. 104, pp. 428–442, Jul. 2019.

[38] K. Raats, V. Fors, and S. Pink, "Trusting autonomous vehicles: An interdisciplinary approach," *Transp. Res. Interdiscipl. Perspectives*, vol. 7, Sep. 2020, Art. no. 100201.

[39] B. C. Choi and A. W. Pak, "Peer reviewed: A catalog of biases in questionnaires," *Preventing Chronic Dis.*, vol. 2, no. 1, 2005, Art. no. A13.

[40] Y. Zhao, X. Wang, M. Goubran, T. Whalen, and E. M. Petriu, "Human emotion and cognition recognition from body language of the head using soft computing techniques," *J. Ambient Intell. Humanized Comput.*, vol. 4, no. 1, pp. 121–140, Feb. 2013.

[41] A. Newen, L. D. Bruin, and S. Gallagher, Eds., *The Oxford Handbook of 4E Cognition. (Oxford Library of Psychology)*. New York, NY, USA: Oxford Univ. Press, 2018.

[42] A. L. Yarbus, “Eye movements during perception of complex objects,” in *Eye Movements Vision*, Berlin, Germany: Springer, 1967, pp. 171–211.
[43] Y. Fang, R. Nakashima, K. Matsumiya, I. Kuriki, and S. Shioiri, “Eye-head coordination for visual cognitive processing,” *Plos One*, vol. 10, no. 3, Mar. 2015, Art. no. e0121035.
[44] F. A. Proudlock, H. Shekhar, and I. Gottlob, “Coordination of eye and head movements during reading,” *Invest. Opthalmology Vis. Sci.*, vol. 44, no. 7, Jul. 2003, Art. no. 2991.
[45] M. F. Land, “The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations,” *Exp. Brain Res.*, vol. 159, no. 2, pp. 151–160, Nov. 2004.
[46] J. F. Grafsgaard, K. E. Boyer, E. N. Wiebe, and J. C. Lester, “Analyzing posture and affect in task-oriented tutoring,” in *Proc. 25th Int. FLAIRS Conf.*, 2012, pp 438–443.
[47] Y. Lu and N. Sarter, “Eye tracking: A process-oriented method for inferring trust in automation as a function of priming and system reliability,” *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 6, pp. 560–568, Dec. 2019.
[48] F. N. Nezami, M. A. Wächter, G. Pipa, and P. König, “Project westdrive: Unity city with self-driving cars and pedestrians for virtual reality studies,” *Front. ICT*, vol. 7, Jan. 2020, Art. no. 1.
[49] K. H. Knuth, “Optimal data-based binning for histograms,” *Digit. Signal Process.*, vol. 95, 2013, Art. no. 102581.
[50] J. Pek, O. Wong, and A. C. M. Wong, “How to address non-normality: A taxonomy of approaches, reviewed, and illustrated,” *Front. Psychol.*, vol. 9, Nov. 2018, Art. no. 2104.
[51] H. Finch and B. French, “A Monte Carlo comparison of robust MANOVA test statistics,” *J. Modern Appl. Stat. Methods*, vol. 12, no. 2, pp. 35–81, Nov. 2013.
[52] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition (Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics)*. New York, NY, USA: Wiley, 1992.
[53] H. Abdi and L. J. Williams, “Tukey’s honestly significant difference (HSD) test,” *Encyclopedia Res. Des.*, vol. 3, no. 1, pp. 1–5, 2010.
[54] C. Ateş, Ö. Kaymaz, H. E. Kale, and M. A. Tekindal, “Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of Type-I error rates,” *Comput. Math. Methods Med.*, vol. 2019, Jul. 2019, Art. no. 2173638.
[55] R. T. Warne, “A primer on multivariate analysis of variance (MANOVA) for behavioral scientists,” *Practical Assessment Res. Eval.*, vol. 19, 2014, Art. no. 19.
[56] R. T. Warne, M. Lazo, T. Ramos, and N. Ritter, “Statistical methods used in gifted education journals, 2006–2010,” *Gifted Child Quart.*, vol. 56, no. 3, pp. 134–149, 2012.
[57] A. Mehrabian, “Communication without words,” in *Communication Theory*, Evanston, IL, USA: Routledge, 2017, pp. 193–200.
[58] H. Gunes and M. Pantic, “Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners,” in *Proc. Int. Conf. Intell. Virtual Agents*, 2010, pp. 371–377.
[59] Z. Hammal, J. F. Cohn, C. Heike, and M. L. Speltz, “What can head and facial movements convey about positive and negative affect?,” in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2015, pp. 281–287.
[60] V. Venkatesh and M. G. Morris, “Why don’t men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior,” *MIS Quart.*, vol. 24, no. 1, pp. 115–139, 2000.
[61] K. Chen and A. H. S. Chan, “A review of technology acceptance by older adults,” *Gerontechnology*, vol. 10, no. 1, pp. 1–12, Jan. 2011.
[62] B. E. Holthausen, P. Wintersberger, B. N. Walker, and A. Riener, “Situational trust scale for automated driving (STS-AD): Development and initial validation,” in *Proc. 12th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2020, pp. 40–47. [Online]. Available: https://doi.org/10.1145/3409120.3410637
[63] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, “Foundations for an empirically determined scale of trust in automated systems,” *Int. J. Cogn. Ergonom.*, vol. 4, no. 1, pp. 53–71, 2000. [Online]. Available: https://doi.org/10.1207/S15327566IJCE0401_04
[64] M. L. van Emmerik, S. C. de Vries, and J. E. Bos, “Internal and external fields of view affect cybersickness,” *Displays*, vol. 32, no. 4, pp. 169–174, 2011.
[65] A. Stephan, “Trust in highly automated driving,” Ph.D. dissertation, Technische Universität Berlin, Berlin, Germany, 2019.