

Effects of Uncertain Trajectory Prediction Visualization in Highly Automated Vehicles on Trust, Situation Awareness, and Cognitive Load

MARK COLLEY, Institute of Media Informatics, Ulm University, Germany

OLIVER SPEIDEL, Institute for Measurement, Control and Microtechnology, Ulm University, Germany

JAN STROHBECK, Institute for Measurement, Control and Microtechnology, Ulm University, Germany

JAN OLE RIXEN, Institute of Media Informatics, Ulm University, Germany

JAN HENRY BELZ, Institute of Media Informatics, Ulm University, Germany

ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

Automated vehicles are expected to improve safety, mobility, and inclusion. User acceptance is required for the successful introduction of this technology. One essential prerequisite for acceptance is appropriately trusting the vehicle's capabilities. System transparency via visualizing internal information could calibrate this trust by enabling the surveillance of the vehicle's detection and prediction capabilities, including its failures. Additionally, concurrently increased situation awareness could improve take-overs in case of emergency. This work reports the results of two online comparative video-based studies on visualizing prediction and maneuver-planning information. Effects on trust, cognitive load, and situation awareness were measured using a simulation (N=280) and state-of-the-art road user prediction and maneuver planning on a pre-recorded real-world video using a real prototype (N=238). Results show that color conveys uncertainty best, that the planned trajectory increased trust, and that the visualization of other predicted trajectories improved perceived safety.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: Visualization Design; Uncertainty Information; Autonomous vehicles.

ACM Reference Format:

Mark Colley, Oliver Speidel, Jan Strohbeck, Jan Ole Rixen, Jan Henry Belz, and Enrico Rukzio. 2023. Effects of Uncertain Trajectory Prediction Visualization in Highly Automated Vehicles on Trust, Situation Awareness, and Cognitive Load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 153 (December 2023), 23 pages. <https://doi.org/10.1145/3631408>

1 INTRODUCTION

Automated vehicles (AVs) are expected to increase traffic safety [31] by removing the human from the driving task, which is often attributed to being one contributing cause (others being, for example, flawed street design) of up to 90% of traffic accidents [75]. Simultaneously, AV users can engage in various non-driving-related activities such as reading, working, or even sleeping [20, 66]. However, we expect first-time users to observe the AV critically. Until full automation, that is, automation that can handle **all possible** situations, is reached, challenges such as under- or overtrust for AV users arise. Regarding undertrust, it was found that most respondents in the USA, Australia, and the UK mentioned at least slight concerns about system failure in unanticipated situations [73].

Authors' addresses: [Mark Colley](mailto:mark.colley@uni-ulm.de), mark.colley@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; [Oliver Speidel](mailto:oliver.speidel@uni-ulm.de), oliver.speidel@uni-ulm.de, Institute for Measurement, Control and Microtechnology, Ulm University, Ulm, Germany; [Jan Strohbeck](mailto:jan.strohbeck@uni-ulm.de), jan.strohbeck@uni-ulm.de, Institute for Measurement, Control and Microtechnology, Ulm University, Ulm, Germany; [Jan Ole Rixen](mailto:jan.ole.rixen@uni-ulm.de), jan.ole.rixen@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; [Jan Henry Belz](mailto:jan.henry.belz@uni-ulm.de), jan.henry.belz@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany; [Enrico Rukzio](mailto:enrico.rukzio@uni-ulm.de), enrico.rukzio@uni-ulm.de, Institute of Media Informatics, Ulm University, Ulm, Germany.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).

2474-9567/2023/12-ART153

<https://doi.org/10.1145/3631408>

This could, in turn, lead to low adoption rates and, consequently, to a lost opportunity regarding increased traffic safety despite most of the respondents desiring to use this technology [73]. In contrast, overtrust, trusting the technology more than the actual capabilities would allow, leads to over-usage and potentially abusive usage of such automated systems. *Calibrated trust* [63], in contrast, refers to a state in which the user’s trust is appropriate to the system’s capabilities.

To support calibrated trust, pedestrian intention [12] was already visualized within the vehicle either on a dashboard or a windshield with three distinct states: crossing, remaining on the sidewalk, and uncertain. However, the effects of visualizing the trajectory prediction of other vehicles and the own planned trajectory as crucial indicators of performance have not yet been investigated. Therefore, we propose to visualize these on a windshield display. Head-Up Displays (HUDs) are already widely available and are expected to be prevalent in future AVs in even larger sizes, for example, for gaming as well [79]. Providing information about the AV’s predictions of other road users and the AV’s planned maneuver could calibrate trust and increase system transparency. While detecting and recognizing relevant objects is the first prerequisite for automated driving functionality [14], assessing the future states of the other mobile actors (pedestrians, cyclists, vehicles) is crucial for one’s own safe trajectory planning. Therefore, numerous works exist with a focus on pedestrian [3, 24, 71, 74] and vehicle trajectory prediction [62, 77]. However, these were not yet evaluated as a method to convey uncertainty.

Contribution statement: (1) An Augmented Reality (AR) visualization technique of uncertain trajectory predictions, (2) findings of an online study based on a simulation ($N=280$) showing that color was most suited to visualize uncertainty, and (3) findings of a second online study ($N=238$) based on a video of a real-world ride with a state-of-the-art vehicle trajectory prediction [77] and maneuver planning [36]. Results show that visualizing the own planned trajectory increased trust in the AV, and visualizing the predictions of other vehicles’ trajectories increased perceived safety. Additionally, the assessment of lateral and longitudinal was increased with the visualizations. This work provides evidence for appropriate communication of AV’s capabilities and intentions and, therefore, helps introduce AVs safely in general traffic.

2 RELATED WORK

Our work focuses on in-vehicle human-machine interfaces and their effect on trust, situation awareness (SA), and cognitive load. However, we also present previous work on trust. Our work is clearly to be distinguished from interfaces targeted towards vulnerable road users despite similar trust-related challenges [18, 19, 23, 44, 45].

2.1 Situation Awareness and Trust in Automated Vehicles

Endsley defined SA as “an understanding of the state of the environment” [28, p. 1] providing the “primary basis for subsequent decision making and performance in the operation of complex, dynamic systems”. Endsley distinguished three levels: Level 1 refers to the recognition of objects. Level 2 refers to conjuncting the objects into an understanding. Level 3, finally, refers to the ability to infer future events based on this understanding. Based on SA, the person can make an informed decision in conjunction with their goals, preconceptions, abilities, complexity, and an available interface. In our case, the visualization of the own and others’ trajectories can help inform the user of the AV about relevant objects (Level 1), the derived future states (Level 3), and the vehicle’s capabilities.

While trust is hard to define [27], a common concept defines trust “as the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” [55, p. 51]. Trust is necessary when (1) there are two agents, (2) there is an incentive, and (3) there is a possibility to fail [55] which is the case in AVs [27]. Ekman et al. [27] defined a framework for AV human-machine interface design. They define the three phases *Pre-Use Phase*, *Learning Phase*, and *Performance Phase*. In each, there is an appropriate level of trust. Is trust too high (i.e., overtrust), then there is misuse. If trust is too low (i.e., undertrust), then there

will be disuse. Several trust-affecting factors were defined, including Mental model, training, anthropomorphism, feedback, why & how information (also see [50]), error information, and **uncertainty information**. According to Ekman et al. [27], this uncertainty information, which is the focus of this work, is most relevant in a control transition from manual to automated driving and **during** the automated journey in the *Learning Phase*, and for a change of context in the *Performance Phase*. In our work, we focus on the uncertainty information via trajectory visualization in the *Learning Phase* as AVs are not yet commercially available.

Detjen et al. [22] also name *Trust & Transparency* one of the acceptance challenge categories for AVs. For trust calibration, the authors encourage uncertainty visualization. While they state that hue is the best visual parameter, they state that suitable design metaphors are “bars, percent values, graphs, or pictograms” [22, p. 11]. They also state that “automation trust depends more on performance-oriented qualities” [22, p. 11]. We argue that these performance-oriented qualities e.g., keeping distances to other road users, recognizing all relevant road users, etc.) could be context-dependent and possibly better visualized with driving-task-related visualizations such as trajectories than with more abstract metaphors such as percent values. Percent values in the automotive context were already argued to be easily misinterpreted or not understood at all [11].

2.2 Visualizing Uncertainty and Effect on Trust

There are two techniques to communicate uncertainty in general: graphical annotations and visual encodings. Figures have five types of uncertainty encodings: fuzziness, location, arrangement, size, and transparency [65], which can be combined. MacEachren et al. [59] additionally includes color (hue, saturation, and value) and shape, orientation, or grain. They report that fuzziness and location work well and that value and arrangement were also rated highly. Size and transparency were deemed potentially usable compared to saturation, which was ranked low [59]. As values and their associated uncertainty value are often mapped independently [21, 49], Value-Suppressing Uncertainty Palettes have been proposed [21]. This allows more values to be communicated with low uncertainty and fewer values with high uncertainty.

Van Der Bles et al. [82] showed that communicating limits of information leads to greater uncertainty by the information receiver. However, this only led to a small decrease in trust.

2.3 In-Vehicle Visualization Concepts

It is an open question which information should be displayed during an automated journey [12, 16, 50, 57, 85, 88].

Two main works conducted interviews regarding the necessity for information presentation. Wintersberger et al. conducted 56 interviews on automated journeys concluding that there is “an inverse correlation between situational trust and participants’ desire for feedback” [88, p. 1]. Relevant information themes are predictability, impact, object characteristics, spatial properties, regulations, and visibility. Wiegand et al. [85] identified 17 relevant situations for information presentation. In a think-aloud study with $N=26$ participants, they found that for unexpected driving behavior, the six main concerns are emotion and evaluation, interpretation and reason, vehicle capability, interaction, future driving prediction, and explanation request times [85].

Regarding the required technology to display such information, using HUDs avoids driver diversion as there is no necessity to look down while providing spatial proximity [35]. While current HUDs are comparatively small (e.g., Volkswagen’s HUD is 217 x 88 mm [1], Mercedes Benz’s is 1.750 x 873 mm large [47]), the entire Windshield Displays (WSDs) are envisioned at continuous depth [38], that is, not only at specific layers (i.e., in certain distances from the vehicle user). Other work investigated various modalities and technologies to communicate with the user. Löcken et al. [58] employed ambient light to communicate the AV’s decisions to the user. van Huysduynen et al. [83] varied ambient light attached to the a-pillars to alter speed perception, generally improving user experience. Wilbrink et al. [86] employed light strips to indicate perceived pedestrians’ intentions or perceptions. Lindemann et al. [56] highlighted potential threats such as pedestrians via an AR WSD. They also

provided a cube over moving vehicles which indicated their behavior type (e.g., dangerous or unusual), resulting in higher SA than only presenting basic information (i.e., *speed* and *navigation info*). While these visualization techniques all showed promising results, the question of which display technology provides the best results was investigated by Colley et al. [16]. They highlight “that no technologically advanced visualization (e.g., AR) is necessary” [16, p. 17].

A related concern is calibrating appropriate trust [12, 14, 16] while providing a good user experience [33, 83]. It was shown that system transparency increases trust in AVs [12, 26, 50]. Kraus et al. [52] provide evidence that high system transparency can even prevent a decrease in the trust after a malfunction. Frison et al. [33] showed that user experience also influences trust in AVs.

Häuslschmid et al. [40] displayed the vehicle’s current situation interpretation via a world in miniature (based on Tesla’s visualizations) or a simulated chauffeur avatar. Trust was increased the most by the world in miniature [40]. Opinions about the necessity of such a visualization varied greatly [40]. von Sawitzky et al. [84] showed the own maneuver or trajectory either on a map, world-fixed or as a combination in a simulated AR WSD. The authors state that they “are aware that even in FAD vehicles will not always be completely reliable and can still fail” [84, p. 1]. However, different from our approach, no uncertainty information is included in their visualizations. von Sawitzky et al. [84] found that such visualizations significantly increase trust. While with a focus on gesture and speech interaction, Tscharn et al. [80] visualized the current traveling path via a blue semi-transparent path. As their focus was on interaction, no insights regarding the visualization can be drawn. Colley et al. [12] compared the visualization of pedestrian intention in a virtual reality (VR) study, finding that the AR version (compared to the tablet version) was significantly better rated in terms of cognitive load. Therefore, we also employed a simulated AR WSD. Using a similar approach, the effects of object detection were evaluated [14]. This increased SA required little cognitive load, and did not increase trust. The authors, therefore, argue that this was a possible visualization technique to calibrate trust. In this approach, the uncertainty of object detection was directly integrated into the visualization. In general, however, automation uncertainty visualization was less investigated. Colley et al. [17] varied the visualization of situation detection, intention, and ego maneuver planning. Kunze et al. [53] showed that especially hue conveys urgency in a sorting study on visual variables. Kunze et al. [54] also argued against using the instrument cluster to reduce workload. Therefore, they used a light strip and a vibrotactile seat, enabling users to focus on the road. Even more abstract representations of uncertainty were used by Beller et al. [4]. They displayed an anthropomorphic symbol at system limits. Results showed that a longer time to collision was available for takeovers. Additionally, SA and trust were higher with the visualization. Helldin et al. [41] used bars to indicate certainty. They also found quicker control takeovers. However, participants trusted the automation less when shown this uncertainty information. Schneider et al. [72] provided live explanations to the passenger of an AV. These included an LED strip and AR information regarding the destination, speed, street name, and planned trajectory. In an unclear situation, the color intensity of the planned trajectory was reduced. The main findings include that a live explanation can reduce negative effects on user experience during an automated journey.

In general, previous work showed the positive influence of communicating the vehicle state and the potential to include uncertainty information. While previous work did include the own and others’ planned trajectories, uncertainty information was not included, nor was a real state-of-the-art algorithm used to deduct these trajectories; instead, these were based on assumptions about how these could work.

2.4 Relation Between Trust, Cognitive Load, and Perceived Safety

According to Hoff and Bashir [43], there are three dimensions to trust: dispositional trust, situational trust, and learned trust. Especially dispositional trust varies concerning internal and external variability. External variability refers to the type of system, system complexity, task difficulty, workload, perceived risks, perceived benefits,

organizational setting, and framing of task (see also Holthausen et al. [46] and Müller et al. [64]). As trust is a psycho-physiological state, it includes cognitive elements. This is because it involves calculating the subjective probability in a specific situation, leading to its correlation with cognitive load [69]. Also, the correlation between cognitive load and trust has been shown in various empirical studies [10, 37, 70].

Therefore, we also measured the constructs cognitive load (refers to workload) and perceived safety (refers to perceived risks and perceived benefits).

Körber [51] based his trust definition on the dimensions by Mayer et al. [60] and Lee and See [55]. In his definition, trust is influenced by Competence / Reliability (see Ability [60] or Performance [55]), Understandability / Predictability (see Integrity [60] or Process [55]), and Intention of Developers (see Benevolence [60] or Purpose [55]). Additional relevant factors are Familiarity and Propensity to Trust. We used the understandability subscale based on his work and his proposed subscale.

3 CONCEPT

We propose to visualize both the results of the trajectory prediction of other vehicles and the planned trajectory to the passenger.

While Colley et al. [12] presented the discrete decision of whether a pedestrian wants to cross with a symbol floating over the person's head, we envision to visualize the future path of other road users. For the AV, only the binary decision of the pedestrian of whether to cross is relevant. Other behavior (e.g., hesitation) ultimately only contributes to the understanding of this binary classification. The estimated trajectory of the other vehicles contains more states: driving on, turning left or right, turning around, stopping. Therefore, visualizing the estimated trajectory is more appropriate than visualizing discrete states. According to Chen et al. [9], mobile road users such as pedestrians, vehicles, and cyclists are the most important traffic objects as these influence the AV's planning. With regards to previous work [12, 14], an AR-based representation seems most appropriate. Compared to center-console visualizations, AR-based ones were found to significantly reduce mental workload [12] and increase trust [12]. Additionally, AR enables information to be presented in perfect spatial relations. Furthermore, the representation should be visualized in all views (i.e., windshield and peripheral or side windows) to provide all relevant information. Additionally, as a user might be unfamiliar with the numerous sensors built into an AV (front, rear, sides), allowing the recognition of objects in the entire surrounding, visualization in all views can support the user to gain a clearer picture of the AV's capabilities. Thus, we expect that showing these predictions in all views to calibrate trust. Displaying the trajectory information allows for the assessment of object detection **and** trajectory prediction as only predictions of detected objects can be made. As the estimation of the own's and the other's trajectory becomes less certain the further the estimation goes into the future, an appropriate form of uncertainty encoding is required. Therefore, we first conducted a simulation-based online survey to find the most appropriate encoding.

4 ONLINE SIMULATION STUDY

4.1 Visual Design

As outlined in Section 2.2 and 2.3, there are numerous possibilities to visualize uncertainty. Based on the contextual nature of the trajectory, we omitted location and orientation variations. We omitted fuzziness to prevent the notion of imperfect visualization (e.g., because of immature technology). We considered size but found in internal tests that this quickly led to visual clutter in scenes with numerous vehicles.

Therefore, based on the work by Kunze et al. [53], Padilla et al. [65], and MacEachren et al. [59], we included the visual variables *color* (hue), *shape*, and *transparency* and adapted these for the use case of trajectory visualization. For the shapes, we employed a *line*, a *double line* representing the vehicle's edges, and a *bezier cone* (see Figure 1). The cone inherently incorporates possible future deviations and, therefore, includes uncertainty information.

While other shapes are possible (e.g., a V-shape), these three shapes represent a prototype for only conveying the direction (line), including information about the size of the AV (double line), and including uncertainty information directly represented in the shape (bezier cone). We also considered other shapes, such as an inverse bezier cone (i.e., a wide start representing certainty ending in a point) but found that this does not align with the possibility of the vehicles to maneuver and provides possibilities for misunderstandings.

The variables *color* and *transparency* had two levels: *with* and *without*. For the color, *blue* to *red* was chosen as done by Kunze et al. [53]. Here, blue at the beginning of the trajectory (near the vehicle) stands for high certainty, and red at the end stands for low certainty. When transparency was included, the shape started solid and “faded out” to convey that the certainty is high near the vehicle (solid) and low the further the trajectory is away from the vehicle (high transparency of the shape).

The visualizations change their length based on the velocity of the own and the other vehicles. This is consistent over all variables. Therefore, the internal consistency for this study is given. The length does alter the perceptibility of the predictions. However, for the study, we chose videos that include a variety of velocities (and, therefore, lengths) to ensure that participants can perceive the predictions. The gradients for color and transparency were adjusted based on the length of the prediction.

4.2 Study Design

Based on the proposed visualizations, we designed the, as a $3 \times 2 \times 2$ design with a baseline. For some of the combinations, see Figure 1. We designed, implemented, and conducted a video-based online between-subject study to evaluate the variables. This exploratory study was guided by the research question:

RQ1 *What impact do the variables “shape”, “color”, and “transparency” have on passengers in an AV in terms of (1) cognitive load, (2) trust, (3) situation awareness, (4) perceived safety, (5) visibility, (6) understandability, and (7) capability assessment?*

4.3 Materials

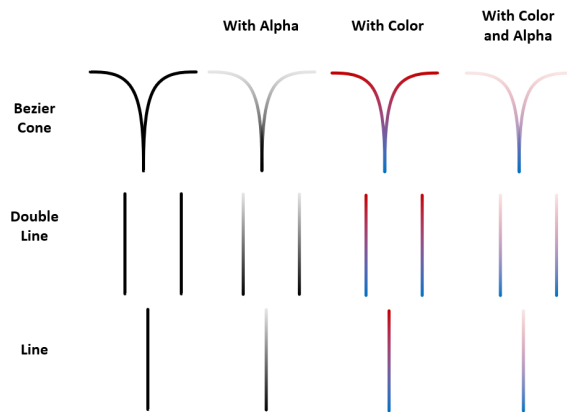


Fig. 1. Overview of the variables shape, color, and transparency. The shapes without color were **white** in the study but are shown in **black** here for better visibility.



Fig. 2. The interior used in the videos for the online study showing the colored line.

For this, we recorded 13 videos of the simulation in Unity [81] (12 visualizations and baseline without any trajectory visualization). The simulation showed the same scene but varied in the described visual properties

of the trajectory visualization. The videos show a ride through a lively city (see Figure 2). Every video took approximately two minutes. Pedestrians and cyclists cross the road both on and without crosswalks. According to Kaß et al. [48], the vehicle performs lateral (i.e., driving straight ahead and turning) and longitudinal (i.e., accelerate and decelerate/break) maneuvers. The future trajectory is deduced based on the known future waypoints and is dependent on the vehicle's velocity: the higher the velocity, the longer the predicted trajectory as route changes become more unlikely, and velocity determines the possible traveled distance for a fixed time frame. Therefore, in the online simulation, the visualization length solely depends on the velocity. To solely assess the trajectory visualization, we also visualized the own trajectory accordingly.

4.4 Measurements

Based on the RQ1, we employed the mental workload subscale of the raw NASA-TLX [39] on a 20-point scale ("How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?"; 1=Very Low to 20=Very High) to measure the cognitive demand of integrating the proposed visualizations into a coherent surroundings assessment. Additionally, we used the subscales *Predictability/Understandability* (*Understanding* from here) and *Trust* of the *Trust in Automation* questionnaire by Körber [51]. Understanding is measured using agreement on four statements ("The system state was always clear to me.", "I was able to understand why things happened."; two inverse: "The system reacts unpredictably.", "It's difficult to identify what the system will do next.") using 5-point Likert scales (1=Strongly disagree to 5=Strongly agree). Trust is measured via agreement on equal 5-point Likert scales on two statements ("I trust the system." and "I can rely on the system."). Also, participants rated their perceived safety using four 7-point semantic differentials from -3 (anxious/agitated/unsafe/timid) to +3 (relaxed/calm/safe/confident) [30]. Regarding SA, we used the SA rating technique (SART) [78]. The SART allowed us to assess the perceived quality of SA [29] which can be viewed as a predictor of "how a person will choose to act on that situation awareness" [29, p. 86]. Depending on the user's capability to understand situations, the user might choose to engage in the driving task and not use the automation. With high perceived qualitative SA, we would expect AV passengers to take over control less and, thus, avoid post-automation effects [6, 61]. Participants also rated the AV's capability of predicting the other vehicle's trajectory ("My vehicle recognized where the other vehicles will drive to very well."), the visibility of the visualization ("The visualization was clearly visible."), and the helpfulness of the visualization ("The visualization helped me understand what the automated vehicle believes the other vehicles will do.") on 7-point Likert scales. Finally, participants were able to provide open feedback.

4.5 Procedure

Every participant was randomly assigned to one of **thirteen** conditions. This represents a $3 \times 2 \times 2$ design (shape, color, and transparency) with a *baseline*. First, participants provided informed consent, answered a demographic questionnaire, and received an overview of the study. Next, the setting was introduced as:

Introduction *You are going to see videos of a highly automated ride through a city. The vehicle automatically steers, brakes, and accelerates (lateral and longitudinal guidance). The automated vehicle will try to predict the other vehicles' future maneuvers (=intention) via its onboard sensors. While watching the video, you are supposed to imagine sitting in such an automated vehicle, follow the entire journey attentively, and then assess it.*

Participants then watched a single video. The scenario was the same (e.g., with the crossing of pedestrians at the same position and time) for all conditions. We explained the visualization for the participant as a combination of the following sentences:

During this ride, your vehicle will [show not show any visualization]/a [line]/[two lines]/[cone] on the street in front of the vehicle, indicating their predicted possible future locations.

The color indicates your automated vehicle's certainty of the prediction of where the other vehicle could be in the future (blue=high certainty to red=low certainty).

The transparency (also known as alpha) level indicates your automated vehicle's certainty of the prediction of where the other vehicle could be in the future (solid=high certainty to transparent= low certainty).

On average, a session lasted 9 min. Participants were compensated with 0.90€. We ran a script in the background to ensure that the window was maximized, that participants were not able to skip or rerun the video (thereby ensuring equal exposure time), and that participants used at least a required Full HD (1980 x 1080 pixels) monitor.

4.6 Results

4.6.1 Data Analysis. Depending on the data (we tested normal distribution and homogeneity of variance assumption), we used the ARTool package by Wobbrock et al. [89]. The procedure is abbreviated, as in the original publication, with ART. In the case of normally distributed data, an ANOVA was used. To compare the systems, we employed the Kruskal-Wallis test for non-normal and a Welch test for normal data. We always used Holm correction for post-hoc tests. Effect sizes are partial Partial eta-squared (η_p^2) calculated with the effectsize package in version 0.8.3 [5]. R in version 4.3.1 and RStudio in version 2023.06.1 were employed. All packages were up to date in July 2023.

4.6.2 Participants. We determined the required sample size via an a-priori power analysis using G*Power in version 3.1.9.7 [32]. To achieve a power of .8 with an alpha level of .05, 273 participants should result in an anticipated medium effect size (0.26 [34]) in an ANOVA.

We recruited, after exclusions due to failed attention checks, 280 participants with an average age of $M=30.73$ ($SD=9.98$) years via <https://www.prolific.co/>. The participant pool was restricted to US citizens to avoid effects of traffic handedness (right-hand vs. left-hand traffic) or culture [67]. While this introduces cultural bias, it also increases internal validity. Using an online participant database allowed us to circumvent biases found when recruiting mostly from a student population (e.g., see almost three-quarters of CHI publications in 2014 [7]). 90 participants identified as male, 189 as female, and one as non-binary. Participants owned a driver's license, on average, for $M=11.68$ ($SD=9.60$) years. Demographic information is shown per condition in Table 1.

Table 1. Number of participants and demographic information per condition. Kruskal-Wallis tests found no significant difference neither for age ($p=0.86$ nor for owning a driver's license ($p=0.97$).

Condition	# participants	Mean age	SD age	Mean license	SD license
Baseline	21	28.67	8.11	9.29	6.72
Bezier Cone Alpha Color	19	32.95	7.52	13.79	8.78
Bezier Cone Alpha	23	30.91	10.59	12.17	10.54
Bezier Color	22	31.68	11.30	13.64	10.91
Bezier Cone	19	33.11	13.62	13.21	12.64
Double Line Alpha Color	23	31.83	9.91	12.09	10.15
Double Line Alpha	22	29.27	7.09	9.82	6.13
Double Line Color	22	30.95	9.21	11.41	9.98
Double Line	21	31.81	9.99	12.90	9.61
Line Alpha Color	24	27.75	11.40	10.00	9.33
Line Alpha	21	30.05	9.15	10.76	8.90
Line Color	22	29.41	8.26	10.64	7.28
Line	21	31.86	12.60	12.64	12.91

101 participants drive daily, 88 on working days, 48 3-4 times a week, 9 one time a week, 7 1-3 times per month, and 27 less than 1 time a month. 63 participants drove less than 7000 km last year, 90 between 7000 and 14999, 69 between 15000 and 24999, 36 between 25000 and 32999, and 22 33000 or more km. Participants indicated that their highest educational level is College (184), followed by High School (48), Vocational Training (32), Middle School (15), and Secondary School (1). Regarding their current employment, participants reported that most of them are employees (145), followed by college students (58), are job seeking (31), are students in school (21), are self-employed (19), or indicated “other” (6). Participants reported medium to high interest in AVs ($M=3.82$, $SD=1.29$), medium belief that AVs will ease their lives ($M=3.76$, $SD=1.24$), and rather believed that AVs would become reality by 2031 ($M=4.18$, $SD=1.00$).

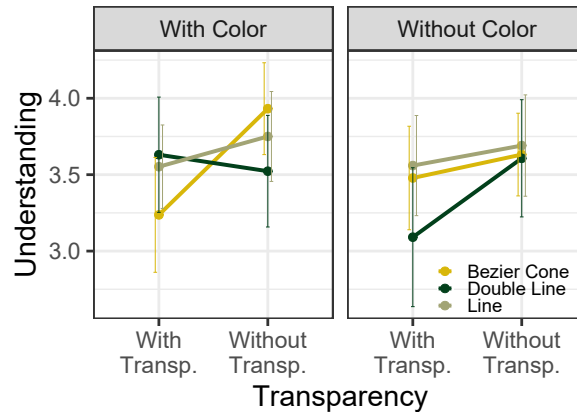


Fig. 3. Three-way interaction effect (IE) on Understanding.

4.6.3 Mental Workload, Trust, and Perceived Safety. The ART found no significant effects on mental workload, which was low in all conditions (between $M=3.09$ (double line with transparency) and $M=3.93$ (Bezier cone with color)).

The ART also found no significant effects on trust or perceived safety.

A Kruskal-Wallis test also found no significant differences for trust ($\chi^2(12)=13.11$, $p=0.36$), perceived safety ($\chi^2(12)=9.59$, $p=0.65$), or mental workload ($\chi^2(12)=8.42$, $p=0.75$).

As the data for Understanding was normally distributed and the homogeneity of variance assumption was met, we calculated the statistical effects using an ANOVA. An ANOVA found a significant effect of transparency on Understanding ($F(1, 247) = 7.57$, $p=0.006$) and a significant three-way IE of transparency on Understanding ($F(2, 247) = 3.09$, $p=0.047$; see Figure 3). Understanding was higher without transparency with all shapes but the Double Line in the colored case.

A Welch-Test found no significant differences in Understanding between conditions ($F(12, 103.3) = 1.57$, $p=0.11$).

4.6.4 Situation Awareness. A Kruskal-Wallis test found no significant differences in SA between conditions ($\chi^2(12)=20.00$, $p=0.07$). The ART found a significant main effect of shape on situation awareness ($F(2, 247) = 3.61$, $p=0.028$, $\eta_p^2 = 0.03$, Cohen’s $f=0.17$). However, pairwise comparisons using Dunn’s test showed the differences not to be significant.

A Kruskal-Wallis test found no significant differences for the SA subscale demand between conditions ($\chi^2(12)=16.20$, $p=0.18$). The ART found a significant IE of color \times transparency on the SA subscale demand

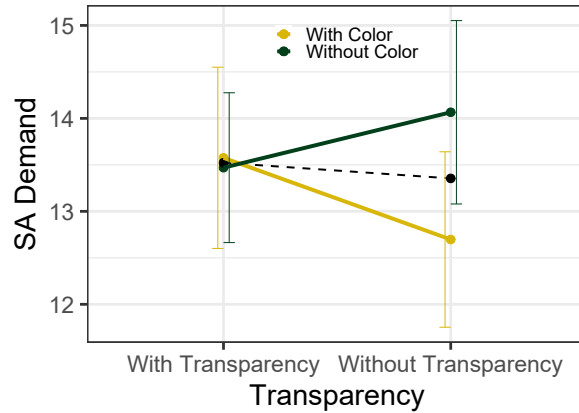


Fig. 4. IE of color \times transparency on Demand.

($F(1, 247) = 4.15, p=0.043, \eta_p^2 = 0.02$, Cohen's $f=0.13$; see Figure 4). While demand was approximately the same with and without color with transparency, without transparency, and with color, the Demand declined, while without color, it increased.

A Kruskal-Wallis test found no significant differences in the SA subscale supply between conditions ($\chi^2(12)=11.49, p=0.49$). The ART found no significant effects on the subscale supply.

A Welch-Test found a significant difference for the subscale Understanding of the SART between conditions ($F(12, 103) = 2.40, p<0.001$). However, post-hoc tests using the Games-Howell test did not reveal significant differences.

An ANOVA found a significant main effect of transparency ($F(1, 247) = 4.18, p=0.042, \eta_p^2 = 0.02$, Cohen's $f=0.13$) on the subscale Understanding of the SART. Understanding was higher without transparency ($M=13.87, SD=3.29$) than with ($M=13.05, SD=3.25$). An ANOVA also found a significant main effect of shape ($F(2, 247) = 6.32, p=0.002, \eta_p^2 = 0.05$, Cohen's $f=0.23$). Post-hoc pairwise t-tests showed that the value for the subscale Understanding of the SART was significantly better for the Line ($M=14.44, SD=2.99$) compared to both the Bezier Cone ($M=12.93, SD=3.18; p_{adj}=0.08$) and the Double Line ($M=12.95, SD=3.48; p_{adj}=0.01$).

4.6.5 Visibility, Prediction Certainty, and Visualization Helpfulness. A Kruskal-Wallis test found no significant differences in visibility between the twelve conditions with a visualization ($\chi^2(11)=5.35, p=0.91$).

The ART found no significant effects on visibility, which was consistently rated very high (from $M=5.74$ (Bezier Cone transparency) to $M=6.52$ (Line)).

A Kruskal-Wallis test found no significant differences in prediction certainty between conditions ($\chi^2(12)=8.25, p=0.77$).

The ART found a significant IE of color \times transparency on prediction certainty ($F(1, 247) = 7.83, p=0.006$). The ART found a significant IE of shape \times color \times transparency on prediction certainty ($F(2, 247) = 3.24, p=0.041$; see Figure 5). Values ranged from $M=5.05$ (Double Line transparency) to $M=6.05$ (Double Line)). Without color, the prediction certainty went up without transparency. With color, it went up for the bezier cone and the line (slightly) but went down for the double line without transparency.

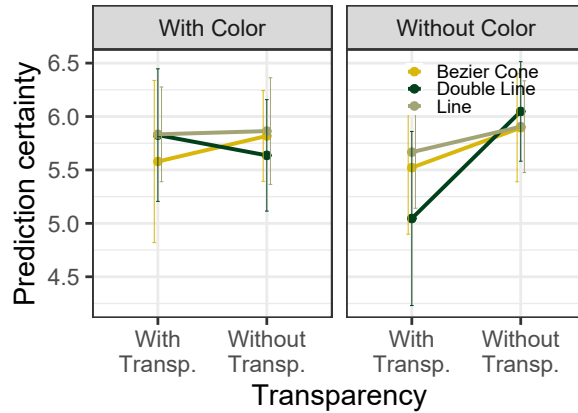


Fig. 5. IE of shape × color × transparency on prediction certainty.

A Kruskal-Wallis test found no significant differences in the helpfulness of the visualization ($p=0.95$; values ranged from $M=5.16$ (Bezier Cone transparency Color) to $M=5.91$ (Bezier Cone Color and Double Line transparency Color)).

4.6.6 Open Feedback. Regarding the visualization, participants agreed that the visualization was relevant and useful. One participant stated, “The visualization was clearly visible. My vehicle recognized where the other vehicles will drive to very well. The visualization helped me understand what the automated vehicle believes the other vehicles will do.” Another participant stated that “those predictions about the vehicles were very helpful”. Regarding the Double Line, which should visualize the vehicle’s edges, one participant stated “The two transparent lines seem to be easily mistakable for headlights”. One participant also wrote “With this line, I feel safer because it shows where the vehicle will go.” Two participants deemed the predictions too short.

4.7 Conclusion

There was no clear best visualization based on the online simulation. Kruskal-Wallis tests found no significant differences between any of the visualizations and the baseline without a visualization. However, based on the open feedback, we assume that a visualization provides aid for the users of AVs. Regarding the design of the visualization, we opted for the colored line. This is based on previous work also using color [53] and the line shape [13, 17, 72]. Understanding was lower with transparency, which is why we omitted this property despite its wide adoption in uncertainty visualization.

For the online simulation study, we used Unity, and we simulated a potential trajectory prediction. In this setting, the prediction was also simulated and does not convey the actual capabilities of today’s state-of-the-art.

5 ONLINE STUDY WITH REALISTIC FOOTAGE

To increase external validity, we additionally used videos taken in the real world and applied state-of-the-art maneuver planning [36] and trajectory prediction [77] for the other vehicles.

While we were generally interested in uncertainty visualizations in the first study and, therefore, always showed both the own planned and predicted trajectories, in this exploratory study, we were interested in the research question:

RQ2 What impact do the visualization of the **own planned maneuver** and the **predicted trajectories of other road users** have on passengers in an AV in terms of (1) cognitive load, (2) trust, (3) situation awareness, (4) perceived safety, (5) visibility, (6) understandability, and (6) capability assessment?

Therefore, the study followed a 2×2 (own planned maneuver visualization: yes/no and visualization of predicted trajectories of other road users: yes/no) between-subjects study design.

5.1 Procedure, Measurements, and Materials

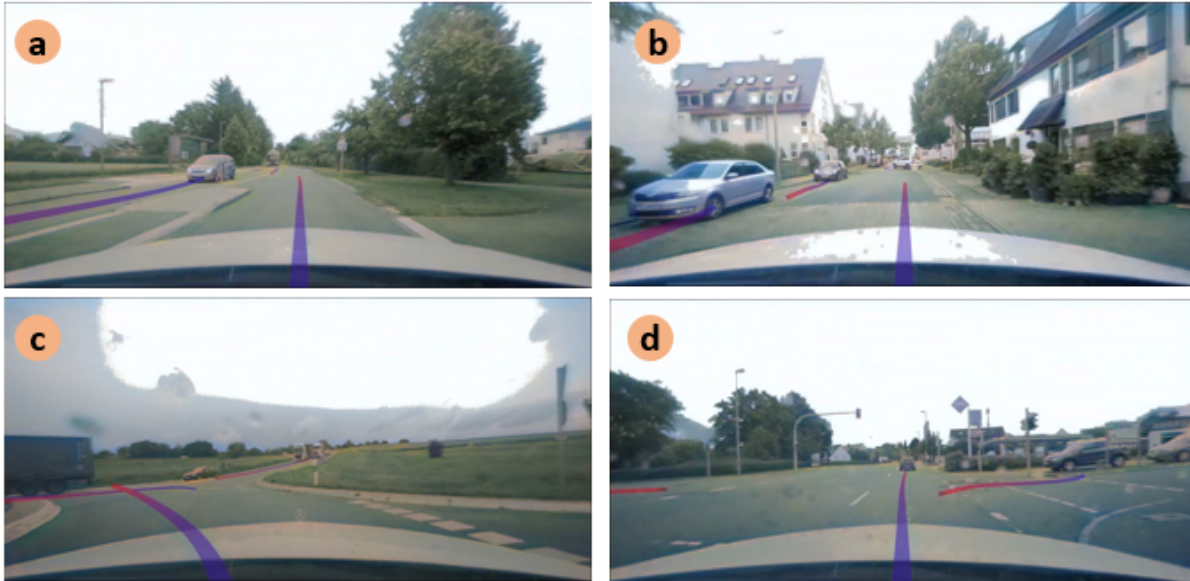


Fig. 6. Screenshots from the colorized videos with the own planned and predicted trajectories. a) shows the exit of a roundabout, b) shows that parked vehicles are also predicted by the algorithms as these could start driving at any moment, c) shows a roundabout, and d) shows an intersection.

We used the same procedure and measurements as in the online simulation study but with only four conditions. We omitted questions regarding visibility but added the subscales *Performance*, *Judgement*, and *Reaction* of the Situational Trust Scale for Automated Driving [46]. Additionally, participants were asked to rate the capability to recognize other vehicles (“The automated vehicle recognizes all vehicles in every situation perfectly”), to predict other vehicle’s paths (“The automated vehicle predicts all vehicle paths in every scene perfectly”), to maneuver longitudinally (“The automated vehicle has perfect longitudinal guidance (braking, acceleration, ...)”) and laterally (“The automated vehicle has perfect lateral guidance (keeping track, staying on the road)”) on 7-point Likert scales (1=Totally Disagree to 7=Totally Agree). These measures were included only in this study as the vehicle in the Online Simulation Study, due to the scripted approach, was able to predict everything perfectly. In this realistic study, a real implementation was used. Thus, some minor uncertainties were to be expected. The videos took approximately 6min and 45s. Participants were compensated with 1.00€.

The introduction to the visualization was given as:

During this ride, your vehicle will show a line on the street in front of the other vehicles indicating their predicted possible future locations. The color indicates your automated vehicle’s certainty of

the prediction of where the other vehicle could be in the future (blue=high certainty to red=low certainty).

The vehicle will show its own planned trajectory. The color indicates your automated vehicle's certainty of the prediction of where it will drive to (blue=high certainty to red=low certainty).

For predicting trajectories of other vehicles, we used the approach described by Strohbeck et al. [77]. The approach uses synthetic agent-centered images to represent the agent's environment to predict, and forecasts multiple hypotheses about the vehicle's future motion over the next few seconds, along with a probability for each hypothesis and uncertainty information along the predicted trajectories. In our case, we trained the neural network to predict six hypotheses for the next three seconds. The prediction network was trained on the Argoverse Motion Forecasting dataset [8], a large-scale dataset of recorded vehicle trajectories with a large variety of > 300000 situations. On the Argoverse Motion Forecasting dataset [8], this approach achieves a minimum Final Displacement Error of 4.19, a Drivable Area Compliance of 0.98, and a Miss Rate of 0.63. It won the Argoverse Motion Forecasting Challenge in 2019. To avoid visual clutter, only the most probable hypothesis was drawn per vehicle.

The trajectory planning framework used in this work follows the approach presented by Graf et al. [36]. The concept is based on the typical two-staged architecture, where first, a behavior trajectory is calculated and subsequently optimized. Thereby, the behavior trajectory is obtained using driver models, which allow for generating reasonable driving behavior efficiently. This behavior trajectory is then integrated into a local continuous optimization problem that considers the comfort of longitudinal and lateral movement, subject to safety and kinematic constraints. The authors report that the "driving experience during this scenario can be described as natural and human-like" [36, p. 129].

We had to use a vehicle equipped with calibrated cameras, radars, and lidars for the approaches to work. The corresponding sensor data is processed to estimate the states of surrounding vehicles using a Labeled Multi-Bernoulli Filter [68]. Together with the ego-state estimation and map data, these form the prediction and trajectory planning module input. As the videos taken are in black and white, we used DeOldify [2] to restore color for the visualization. The final result can be seen in Figure 6. While the vehicle would have been capable of driving automated, we drove manually to relieve some of the processing power for a more reliable data recording. As a result, the predictions and the planned maneuver were calculated after the ride but are real-time capable. That means that, for the experiment, there was no possibility of a delay in the prediction. Potential inaccuracies of the systems for this experiment are not to be seen as an error but are part of the uncertainty in the prediction visualization.

5.2 Results

5.2.1 Participants. To achieve a power of .8 with an alpha level of .05, 232 participants should result in an anticipated medium effect size (0.22 [34]) in an ANOVA. Therefore, we recruited 362 participants via <https://www.prolific.co/>. The participant pool was again restricted to US citizens to avoid the effects of traffic handedness (right-hand vs. left-hand traffic). After discarding participants with failed attention checks, we used 238 data sets for the analysis (54 in the baseline, 62 where only the maneuver was displayed, 61 when only the other vehicle's prediction was shown, and 61 when both were shown). The participants were not identical to the *Online Simulation Study*.

On average, participants were $M=35.87$ ($SD=13.86$; range: 18 – 82) years old. 121 participants identified as female, 115 as male, and two as non-binary. Participants owned a driver's license, on average, for $M=16.63$ ($SD=14.42$) years. The demographic information is shown per condition in Table 2.

129 participants drive daily, 35 on working days, 41 3-4 times a week, 10 one time a week, 9 1-3 times per month, and 14 less than one time a month. 30 participants drove less than 7000 km last year, 26 between 7000 and

Table 2. Number of participants and demographic information per condition. Kruskal-Wallis tests found no significant difference neither for age ($p=0.28$) nor for owning a driver's license ($p=0.76$)

Condition	# participants	Mean age	SD age	Mean license	SD license
Baseline	54	35.26	13.01	16.04	13.45
Only Ego	62	36.89	13.51	17.02	13.95
Only Predictions	61	37.48	14.79	17.98	15.83
Ego and Predictions	61	33.75	14.00	15.43	14.48

14999, 49 between 15000 and 24999, 26 between 25000 and 32999, and 25 33000 or more km. Participants indicated that their highest educational level is College (193), followed by High School (32) and Vocational Training (13). Regarding their current employment, participants reported that most of them are employees (149), followed by college students (29), are self-employed (22), are seeking a job (15), indicated "other" (22), or are students in school (1). On a 5-point Likert scale (1=strongly disagree, 5=strongly agree), participants reported medium to high interest in AVs ($M=3.92$, $SD=1.17$), medium belief that AVs will ease their lives ($M=3.91$, $SD=1.16$), and rather believed that AVs would become reality by 2031 ($M=4.11$, $SD=1.01$).

5.2.2 Mental Workload, Trust, and Perceived Safety. Neither a Kruskal-Wallis test ($p=0.73$) nor the ART found significant effects on mental workload.

Also, neither a Kruskal-Wallis test ($p=0.27$) nor the ART found significant effects on understanding.

A Kruskal-Wallis test found no significant differences in trust ($p=0.09$). The ART found a significant main effect of visualization of the own planned maneuver on trust ($F(1, 152) = 8.61$, $p=0.004$, $\eta_p^2 = 0.03$, Cohen's $f=0.17$). Trust was significantly higher with ($M=3.71$, $SD=0.91$) than without ($M=3.41$, $SD=1.00$) the visualization of the own planned trajectory.

A Kruskal-Wallis test ($\chi^2(3)=7.67$, $p=0.05$) and the ART ($p=0.05$ for visualization of the own planned trajectories) found no significant differences in perceived safety.

5.2.3 Situation Awareness and Capability Assessment. Kruskal-Wallis tests found no significant effects neither on SA ($\chi^2(3)=2.00$, $p=0.57$) nor on the subscales demand ($\chi^2(3)=3.09$, $p=0.38$), or understanding ($\chi^2(3)=5.17$, $p=0.16$). A Welch-test found no significant difference for the SA subscale supply ($F(3, 82.24) = 0.37$, $p=0.77$)

An ART also found no significant effects on any of the scales but Understanding. The ART found a significant main effect of visualization of the own planned trajectories on Understanding ($F(1, 234) = 4.07$, $p=0.045$, $\eta_p^2 = 0.02$, Cohen's $f=0.13$). Understanding was higher with ($M=14.57$, $SD=3.51$) than without ($M=13.76$, $SD=3.63$) the ego trajectory.

A Kruskal-Wallis test found no significant effects neither on the perceived ability to recognize other vehicles ($\chi^2(3)=1.83$, $p=0.61$) nor on the perceived ability to predict other vehicles' trajectory ($\chi^2(3)=1.11$, $p=0.77$). An ART also found no significant effects.

Kruskal-Wallis tests found significant effects on perceived longitudinal ($\chi^2(3)=14.53$, $p<0.001$) and lateral control ($\chi^2(3)=16.44$, $p<0.001$). Post-hoc tests using Dunn's test showed in both cases that the *baseline* ($M=4.88$ for longitudinal and $M=5.14$ for lateral) was rated significantly worse than the prediction-only visualization ($M=5.62$ for longitudinal and $M=6.03$ for lateral) and the combined visualization ($M=5.83$ for longitudinal and $M=5.98$ for lateral).

The ART found a significant main effect visualization of predicted trajectories ($F(1, 234) = 8.38$, $p=0.004$, $\eta_p^2 = 0.03$, Cohen's $f=0.19$) on longitudinal control. The visualization ($M=5.67$, $SD=1.30$) to higher scores (without: $M=5.24$, $SD=1.30$).

The ART found a significant main effect of visualization of predicted trajectories on lateral control ($F(1, 234) = 10.42$, $p=0.001$, $\eta_p^2 = 0.04$, Cohen's $f=0.21$). The ART also found a significant IE of visualization of the own planned

trajectories \times visualization of predicted trajectories on lateral control ($F(1, 234) = 6.47, p=0.012, \eta_p^2 = 0.03$, Cohen's $f=0.17$; see Figure 7). Lateral control was rated high with visualization of predicted trajectories ($M=5.93, SD=1.18$) and lower without the visualization of predicted trajectories ($M=5.51, SD=1.17$). However, the decline was smaller with the visualization of the own planned trajectories.

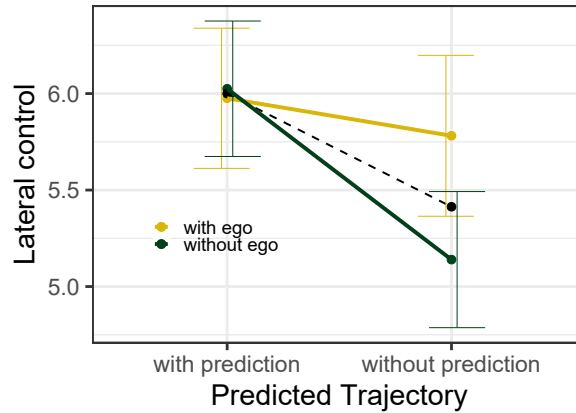


Fig. 7. IE of visualization of the own planned trajectories \times visualization of predicted trajectories on lateral control.

5.2.4 Judgment, Performance, Reaction, Clarity. Kruskal-Wallis tests found no significant differences between conditions for Judgement ($\chi^2(3)=1.77, p=0.62$), Performance ($\chi^2(3)=3.75, p=0.29$), Reaction ($\chi^2(3)=5.71, p=0.13$), or clarity of the next actions ($\chi^2(3)=6.96, p=0.07$).

ARTs also found no significant effects on Performance and Reaction. The ART found a significant main effect of visualization of predicted trajectories on Judgement ($F(1, 234) = 8.34, p=0.004, \eta_p^2 = 0.03$, Cohen's $f=0.19$). Without the visualization of predicted trajectories ($M=2.44, SD=1.47$), Judgement was rated higher than with $M=2.30, SD=1.60$). An ART found, however, a significant difference in the clarity of the next actions ($F(1, 234) = 5.62, p=0.019, \eta_p^2 = 0.02$, Cohen's $f=0.15$) based on visualization of the own planned trajectories. With visualization of the own planned trajectories ($M=5.10, SD=1.52$), it was significantly clearer what the AV will do than without ($M=4.90, SD=1.49$). The ART found a significant IE of visualization of the own planned trajectories \times visualization of predicted trajectories on clarity of the next actions ($F(1, 234) = 4.31, p=0.039, \eta_p^2 = 0.02$, Cohen's $f=0.14$; see Figure 8). While clarity was almost equal with the ego trajectory independent of the availability of predicted other trajectories, this clarity was much lower when the ego was gone and there was no prediction.

5.2.5 Open Feedback. Feedback regarding the visualization was mainly positive. Most participants stated that the visualizations helped them understand the AV's actions (e.g., "The video was really neat to see. I've often wondered what it would be like to be in a self-driving car and had a lot of doubts. I'm still not convinced but I'm more open to it then ever before."). However, one participant was more critical of the visualization of the own planned maneuver. The participant stated that the visualization with the uncertainty shown via the color made them doubt whether the AV could perform the driving task. Another participant proposed a cone-shaped visualization to better visualize the potential space occupied by the other vehicles. Various participants also indicated their astonishment at how well the predictions and the driving style are.

When not having the visualization, one participant stated: "Not knowing where the vehicle is going and me worried the whole time so i think that might be an issue". There were also additional comments to situations not

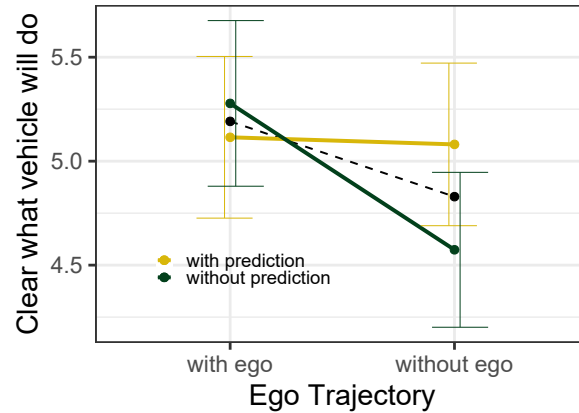


Fig. 8. IE of visualization of the own planned trajectories \times visualization of predicted trajectories on clarity of the next actions.

shown in the videos, such as: “What happens if the cameras become covered with ice?” or “I am not sure how the vehicle would react in situations in which the right of way is complicated. I would not feel safe passengering in an automated car if, for example, we were turning left at a light with a green circle and had to yield to oncoming traffic” or “I am wondering what the car will do in a situation where it comes upon a police officer directing traffic. Can it respond appropriately to the officer’s hand signals?” or “I think to properly evaluate it, I would want to see heavy traffic, things darting out in front of it, etc.”

6 DISCUSSION

This work presents the results of two online studies with $N=280$ and $N=238$ participants. In the first study, we evaluated the most appropriate trajectory planning and prediction visualization under uncertain circumstances. In line with previous work by Kunze et al. [53], we deemed the color to convey uncertainty best. Therefore, in the subsequent study using real-world videos and state-of-the-art algorithms, we evaluated the effect of visualizing the own planned maneuver and/or the predicted trajectories of the other vehicles.

6.1 Calibration of Passenger Expectations

Schneider et al. [72] provided evidence that a live explanation during an automated journey improves user experience. Our data support that showing the own planned trajectory, for example, improves trust significantly. Therefore, such explanatory approaches seem to have multiple beneficial outcomes. While the own planned trajectory implicitly includes the assumptions about the other vehicles as these have to be taken into account [9], we explicitly evaluated the effect of displaying these visualizations. The data showed that perceived safety was higher with the visualized predicted trajectories. We found no significant effects on the SA or the capability assessment. However, the *baseline* was rated significantly worse in lateral and longitudinal control. Our results are somewhat contrary to the work by Colley et al. [14], who showed that visualizing the semantic segmentation task did not lead to increased trust but increased capability assessment. The reasons for this are not clear. The scene (busy inner-city vs. calm town) and the abstraction of the information (low abstraction in the semantic segmentation task vs. higher abstraction for the prediction and planning) could be an explanation. Nonetheless, the open feedback of the participants suggests that the visualizations were useful for them to understand the capabilities of the AV. We assume that visualizing the uncertainty of the planning and prediction is crucial to

avoid overtrust in an AV. Therefore, we argue that the more explicit use of color (compared to color intensity [72]) should be focused on. Therefore, we argue that the chosen visualizations are appropriate to calibrate passengers' expectations of AVs. The open feedback suggests that this might make some user "nervous". While this is an unwanted effect for manufacturers, this, depending on the prior beliefs of the user, is a desired property of the visualization. Assessing the appropriate trust level, however, is an unsolved challenge.

6.2 On the Visualization Design

The proposed visualizations had no significant effect on mental workload. Therefore, we argue that the visualizations are an appropriate tool to convey information about the own planned and predicted trajectories. The design using a color gradient from blue to red was based on the online simulation study showing that color helped in Understanding (see Figure 3) and lowered demand (see Figure 4). While the participants assessed the visualizations as helpful, chevrons could improve understandability by providing directional information. Additional elements, such as stopping lines, could further improve the user experience. Our simulation study showed that transparency (in line with previous work [53]) was inappropriate in conveying uncertainty information. We decided against discrete states (e.g., turning, stopping) as the continuous nature of the line allows us to continuously assess the correctness of the prediction and the own planned maneuver. However, participants requested this kind of information in their open feedback. Furthermore, our work showed that even without providing explanations for the own trajectory, visualizing it increased trust significantly.

6.3 On the Study Methodology

In both experiments, we employed a video-based methodology with simulated and recorded situations. Therefore, the setting of the study limited the possibility of simulating the AV's vection. Future work could improve this aspect by using VR (for a higher immersion) and simulators with higher degrees of freedom (e.g., [15, 42]). Another drawback is that the real-world videos had to be captured in black and white and were, therefore, recolored. Thus, while the result is convincing, informal discussions indicate that the quality is lower than Full HD RGB videos. We see three reasons why the approach is still valid. (1) A real-world scenario was not possible due to the unavailability of AVs and WSDs, as well as the requirement to run advanced algorithms for intention prediction. While it would have been possible to study the effects of these visualizations, for example, in VR, we opted for the video-based methodology due to the increased diversity of participants and the possibility of recruiting sufficiently large numbers of participants, as done in other work [14, 72]. (2) Prior work in which the experiences of passengers in an AV at an intersection with crossing pedestrians was evaluated in a monitor-based setup, a simulator, and in an actual vehicle showed that most (but not all) effects could be found with the monitor-based setup [76]. (3) As it is unclear whether a manufacturer will include such visualizations in their final design, these visualizations could be used to educate the public about the potential drawbacks of this technology and, thus, to calibrate their trust. As such an education could be most effectively via a website or, for example, YouTube videos, we argue that the effects evaluated in the two surveys are relevant to inform future work.

6.4 Practical Implications

The study using the real-world videos confirmed some of the results of related work that transparent systems [26, 50] and AR systems [12, 14, 16, 40, 87] lead to higher trust and perceived safety. Already, HUDs are available in commercially available vehicles (e.g., [1]). These HUDs already allow displaying at least the own planned trajectory. Therefore, we argue that our data shows that the introduction of AVs should be accompanied by the visualization of driving-task relevant information. While AR WSD will be beneficial also to visualize information regarding other road users, the already available technology should be used for trust enhancement and calibration. If manufacturers

do not want to incorporate these visualizations due to marketing concerns (see “autonowashing” [25]), then a different approach has to be found to educate the public about current AV capabilities.

6.5 Limitations and Future Work

In the two studies, numerous participants took part ($N=280$ and $N=238$). However, the demographic information shows that especially younger people participated. This demographic can be attributed to the younger demography on prolific.co. It is, therefore, unclear whether the findings of our two studies can be transferred to other demographic groups (i.e., age groups). Furthermore, due to the setting of the study as an online video-based study, we were only able to measure and study subjective dependent measures such as trust. Therefore, while AVs are not generally available, the results of this study have to be confirmed in real-world settings when possible. This setting also limits the generalizability of the findings to real-world driving conditions.

As we also drove manually during the video recording for the study with realistic footage, the driving style was actually not determined by the algorithm. While this introduces a potential deviation between the projection of the maneuver and the actual ride, we believe this to be minor. Also, no participant mentioned this.

There might also be a contortion in the data by inattentive participants (e.g., for mental workload) despite our best efforts to conduct an internally and externally valid study, for example, by including attention checks.

Finally, participants self-rated their understanding of the visualizations. However, this could have been erroneous, which could have led to misconstrual errors. However, the qualitative results showed no such pattern and the included attention checks should most likely have reduced mischievous participants.

Future work should also consider the effects of such uncertainty visualizations on usability and user experience measures. These could also influence trust. Additionally, personality characteristics and personal values also could influence trust, therefore, future work is encouraged in this area. Additionally, the open feedback showed the necessity to evaluate the visualizations in other scenarios and also the possibility to add more discrete visualizations (e.g., upcoming actions instead of continuous paths). Finally, our work shows the effects of different visualization patterns and the effect of visualizing one pre-trained approach. To actually evaluate calibration, a quantification of automation capability has to be defined and different capability levels must be compared. This is possible with the used methodology but requires future work.

7 CONCLUSION

Overall, this work showed the potential to visualize the own planned trajectory and the predicted trajectory of other vehicles. These tasks are crucial for the successful operation of AVs. In the first online simulation-based study ($N=280$), we evaluated the use of shape, transparency, and color to convey uncertainty. Afterward, we used color to convey uncertainty in visualizing the planned and predicted trajectories using state-of-the-art real-world footage approaches, increasing external validity. Finally, we evaluated the effects of showing the own planned and the other predicted trajectories in an online video-based study ($N=238$), finding that the visualizations increased trust and perceived safety. Our work further shows possibilities to convey information about AVs for their successful introduction.

ACKNOWLEDGMENTS

We thank Sebastian Hartwig for his support regarding the re-colorization of the videos and all study participants. This work was partially supported by the project ‘SEMULIN’ (**s**elbstunterstützende, **m**ultimodale **I**nteraktion) funded by the Federal Ministry for Economic Affairs and Energy (BMWi).

REFERENCES

- [1] Volkswagen AG. 2020. Head-up-Display. <https://www.volkswagen-newsroom.com/de/head-up-display-3957>. [Online; accessed: 07-AUGUST-2021].

- [2] Jason Antic. 2021. DeOldify. <https://github.com/jantic/DeOldify>. [Online; accessed: 05-AUGUST-2021].
- [3] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. 2013. Intention-Aware Motion Planning. In *Algorithmic Foundations of Robotics X*, Emilio Frazzoli, Tomas Lozano-Perez, Nicholas Roy, and Daniela Rus (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 475–491.
- [4] Johannes Beller, Matthias Heesen, and Mark Vollrath. 2013. Improving the Driver–Automation Interaction: An Approach Using Automation Uncertainty. *Human Factors* 55, 6 (2013), 1130–1141. <https://doi.org/10.1177/0018720813482327> arXiv:<https://doi.org/10.1177/0018720813482327> PMID: 24745204.
- [5] Mattan S. Ben-Shachar, Daniel Lüdecke, and Dominique Makowski. 2020. effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software* 5, 56 (2020), 2815. <https://doi.org/10.21105/joss.02815>
- [6] S. Brandenburg and E. M. Skottke. 2014. Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving?. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, New York, NY, USA, 2978–2983. <https://doi.org/10.1109/ITSC.2014.6958168>
- [7] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 981–992. <https://doi.org/10.1145/2858036.2858498>
- [8] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. 2019. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New York, NY, USA, 8748–8757.
- [9] Bike Chen, Chen Gong, and Jian Yang. 2018. Importance-aware semantic segmentation for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems* 20, 1 (2018), 137–148. <https://doi.org/10.1109/TITS.2018.2801309>
- [10] Fang Chen. 2013. *Effects of cognitive load on trust*. Technical Report. NATIONAL INFORMATION COMMUNICATION TECHNOLOGY AUSTRALIA LTD EVELEIGH.
- [11] Mark Colley, Ali Askari, Marcel Walch, Marcel Woide, and Enrico Rukzio. 2021. ORIAS: On-The-Fly Object Identification and Action Selection for Highly Automated Vehicles. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Association for Computing Machinery, New York, NY, USA, 79–89. <https://doi.org/10.1145/3409118.3475134>
- [12] Mark Colley, Christian Bräuner, Mirjam Lanzer, Marcel Walch, Martin Baumann, and Enrico Rukzio. 2020. Effect of Visualization of Pedestrian Intention Recognition on Trust and Cognitive Load. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) (*AutomotiveUI '20*). ACM, Association for Computing Machinery, New York, NY, USA, 181–191. <https://doi.org/10.1145/3409120.3410648>
- [13] Mark Colley, Julian Britten, Simon Demharter, Tolga Hisir, and Enrico Rukzio. 2022. Feedback Strategies for Crowded Intersections in Automated Traffic – A Desirable Future?. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) (*AutomotiveUI '22*). Association for Computing Machinery, New York, NY, USA, 243–252. <https://doi.org/10.1145/3543174.3545255>
- [14] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21). Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445351>
- [15] Mark Colley, Pascal Jansen, Enrico Rukzio, and Jan Gugenheimer. 2022. SwiVR-Car-Seat: Exploring Vehicle Motion Effects on Interaction Quality in Virtual Reality Automated Driving Using a Motorized Swivel Seat. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 150 (dec 2022), 26 pages. <https://doi.org/10.1145/3494968>
- [16] Mark Colley, Svenja Krauß, Mirjam Lanzer, and Enrico Rukzio. 2021. How Should Automated Vehicles Communicate Critical Situations? A Comparative Analysis of Visualization Concepts. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 94 (2021), 23 pages. <https://doi.org/10.1145/3478111>
- [17] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of Scene Detection, Scene Prediction, and Maneuver Planning Visualizations on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 49 (jul 2022), 21 pages. <https://doi.org/10.1145/3534609>
- [18] Mark Colley and Enrico Rukzio. 2020. A Design Space for External Communication of Autonomous Vehicles. In *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (*AutomotiveUI '20*). ACM, Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3409120.3410646>
- [19] Mark Colley, Marcel Walch, Jan Gugenheimer, and Enrico Rukzio. 2019. Including People with Impairments from the Start: External Communication of Autonomous Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) (*AutomotiveUI '19*). Association for Computing Machinery, New York, NY, USA, 307–314. <https://doi.org/10.1145/3349263.3351521>
- [20] Mark Colley, Dennis Wolf, Sabrina Böhm, Tobias Lahmann, Luca Porta, and Enrico Rukzio. 2021. Resync: Towards Transferring Somnolent Passengers to Consciousness. In *Adjunct Publication of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) (*MobileHCI '21*). Association for Computing Machinery, New York, NY, USA, Article 1, 6 pages.

- <https://doi.org/10.1145/3447527.3474847>
- [21] Michael Correll, Dominik Moritz, and Jeffrey Heer. 2018. Value-Suppressing Uncertainty Palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3173574.3174216>
 - [22] Henrik Detjen, Sarah Faltaous, Bastian Pflöging, Stefan Geisler, and Stefan Schneegass. 2021. How to increase automated vehicles' acceptance through in-vehicle interaction design: A review. *International Journal of Human-Computer Interaction* 37, 4 (2021), 308–330. <https://doi.org/10.1080/10447318.2020.1860517> arXiv:<https://doi.org/10.1080/10447318.2020.1860517>
 - [23] 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* (Toronto, ON, Canada) (*AutomotiveUI '18*). Association for Computing Machinery, New York, NY, USA, 82–86. <https://doi.org/10.1145/3239092.3265946>
 - [24] Frederik Diederichs, Tobias Schüttke, and Dieter Spath. 2015. Driver intention algorithm for pedestrian protection and automated emergency braking systems. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, IEEE, New York, NY, USA, 1049–1054.
 - [25] Liza Dixon. 2020. Autonowashing: The greenwashing of vehicle automation. *Transportation research interdisciplinary perspectives* 5 (2020), 100113.
 - [26] Na Du, Jacob Haspiel, Qiaoning Zhang, Dawn Tilbury, Anuj K. Pradhan, X. Jessie Yang, and Lionel P. Robert. 2019. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies* 104 (2019), 428–442. <https://doi.org/10.1016/j.trc.2019.05.025> ID: 271729.
 - [27] Fredrick Ekman, Mikael Johansson, and Jana Sochor. 2018. Creating Appropriate Trust in Automated Vehicle Systems: A Framework for HMI Design. *IEEE Transactions on Human-Machine Systems* 48, 1 (2018), 95–101. <https://doi.org/10.1109/THMS.2017.2776209>
 - [28] Mica R Endsley. 1995. Measurement of situation awareness in dynamic systems. *Human factors* 37, 1 (1995), 65–84.
 - [29] Mica R Endsley, Stephen J Selcon, Thomas D Hardiman, and Darryl G Croft. 1998. A comparative analysis of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 42. SAGE Publications Sage CA: Los Angeles, CA, SAGE Publications, Los Angeles, CA, USA, 82–86.
 - [30] Stefanie M. Faas, Andrea C. Kao, and Martin Baumann. 2020. A Longitudinal Video Study on Communicating Status and Intent for Self-Driving Vehicle – Pedestrian Interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376484>
 - [31] Daniel J Fagnant and Kara Kockelman. 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77 (2015), 167–181.
 - [32] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160.
 - [33] Anna-Katharina Frison, Philipp Wintersberger, Andreas Rienen, 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* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300374>
 - [34] David C. Funder and Daniel J. Ozer. 2019. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science* 2, 2 (2019), 156–168. <https://doi.org/10.1177/2515245919847202> arXiv:<https://doi.org/10.1177/2515245919847202>
 - [35] J. L. Gabbard, G. M. Fitch, and H. Kim. 2014. Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications. *Proc. IEEE* 102, 2 (2014), 124–136.
 - [36] Maximilian Josef Graf, Oliver Michael Speidel, Jona Ruoff, and Klaus Dietmayer. 2021. On-Road Motion Planning for Automated Vehicles at Ulm University. , 2-12 pages. <https://doi.org/10.1109/MITS.2021.3084534>
 - [37] Kunal Gupta, Ryo Hajika, Yun Suen Pai, Andreas Duenser, Martin Lochner, and Mark Billinghurst. 2019. In AI We Trust: Investigating the Relationship between Biosignals, Trust and Cognitive Load in VR. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology* (Parramatta, NSW, Australia) (*VRST '19*). Association for Computing Machinery, New York, NY, USA, Article 33, 10 pages. <https://doi.org/10.1145/3359996.3364276>
 - [38] Renate Häuslschmid, Yixin Shou, John O'Donovan, Gary Burnett, and Andreas Butz. 2016. First Steps towards a View Management Concept for Large-Sized Head-up Displays with Continuous Depth. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Ann Arbor, MI, USA) (*AutomotiveUI '16*). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3003715.3005418>
 - [39] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, The Netherlands, 139–183.
 - [40] Renate Häuslschmid, Max von Bülow, Bastian Pflöging, and Andreas Butz. 2017. Supporting Trust in Autonomous Driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) (*IUI '17*). Association for Computing Machinery,

- New York, NY, USA, 319–329. <https://doi.org/10.1145/3025171.3025198>
- [41] Tove Helldin, Göran Falkman, Maria Riveiro, and Staffan Davidsson. 2013. Presenting System Uncertainty in Automotive UIs for Supporting Trust Calibration in Autonomous Driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Eindhoven, Netherlands) (*AutomotiveUI '13*). Association for Computing Machinery, New York, NY, USA, 210–217. <https://doi.org/10.1145/2516540.2516554>
- [42] Philipp Hock, Mark Colley, Ali Askari, Tobias Wagner, Martin Baumann, and Enrico Rukzio. 2022. Introducing VAMPIRE – Using Kinaesthetic Feedback in Virtual Reality for Automated Driving Experiments. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) (*AutomotiveUI '22*). Association for Computing Machinery, New York, NY, USA, 204–214. <https://doi.org/10.1145/3543174.3545252>
- [43] 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.
- [44] Kai Holländer, Mark Colley, Enrico Rukzio, and Andreas Butz. 2021. A Taxonomy of Vulnerable Road Users for HCI Based On A Systematic Literature Review. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 158, 13 pages. <https://doi.org/10.1145/3411764.3445480>
- [45] Kai Holländer, Philipp Wintersberger, and Andreas Butz. 2019. Overtrust in External Cues of Automated Vehicles: An Experimental Investigation. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) (*AutomotiveUI '19*). Association for Computing Machinery, New York, NY, USA, 211–221. <https://doi.org/10.1145/3342197.3344528>
- [46] Brittany E. Holthausen, Philipp Wintersberger, Bruce N. Walker, and Andreas Riener. 2020. Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) (*AutomotiveUI '20*). Association for Computing Machinery, New York, NY, USA, 40–47. <https://doi.org/10.1145/3409120.3410637>
- [47] Markus Jordan. 2020. DAS MBUX AUGMENTED REALITY HEAD-UP DISPLAY DER S-KLASSE. <https://mbpassion.de/2020/09/das-mbux-augmented-reality-head-up-display/>. [Online; accessed 01-APRIL-2022].
- [48] Christina Kaß, Stefanie Schoch, Frederik Naujoks, Sebastian Hergeth, Andreas Keinath, and Alexandra Neukum. 2020. Standardized Test Procedure for External Human–Machine Interfaces of Automated Vehicles. *Information* 11, 3 (2020), 173.
- [49] Matthew Kay. 2019. How Much Value Should an Uncertainty Palette Suppress if an Uncertainty Palette Should Suppress Value? Statistical and Perceptual Perspectives. , 4 pages.
- [50] 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 (IJDeM)* 9, 4 (2015), 269–275.
- [51] Moritz Körber. 2019. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 13–30.
- [52] Johannes Kraus, David Scholz, Dina Stiegemeier, and Martin Baumann. 2020. The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency. *Human Factors* 62, 5 (2020), 0018720819853686. <https://doi.org/10.1177/0018720819853686> arXiv:<https://doi.org/10.1177/0018720819853686> PMID: 31233695.
- [53] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2018. Augmented Reality Displays for Communicating Uncertainty Information in Automated Driving. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) (*AutomotiveUI '18*). Association for Computing Machinery, New York, NY, USA, 164–175. <https://doi.org/10.1145/3239060.3239074>
- [54] Alexander Kunze, Stephen J. Summerskill, Russell Marshall, and Ashleigh J. Filtness. 2019. Conveying Uncertainties Using Peripheral Awareness Displays in the Context of Automated Driving. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) (*AutomotiveUI '19*). Association for Computing Machinery, New York, NY, USA, 329–341. <https://doi.org/10.1145/3342197.3344537>
- [55] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.
- [56] Patrick Lindemann, Tae-Young Lee, and Gerhard Rigoll. 2018. Catch my drift: Elevating situation awareness for highly automated driving with an explanatory windshield display user interface. *Multimodal Technologies and Interaction* 2, 4 (2018), 71.
- [57] Andreas Löcken, Anna-Katharina Frison, Vanessa Fahn, Dominik Kreppold, Maximilian Götz, and Andreas Riener. 2020. Increasing User Experience and Trust in Automated Vehicles via an Ambient Light Display. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) (*MobileHCI '20*). Association for Computing Machinery, New York, NY, USA, Article 38, 10 pages. <https://doi.org/10.1145/3379503.3403567>
- [58] Andreas Löcken, Wilko Heuten, and Susanne Boll. 2016. AutoAmbiCar: Using Ambient Light to Inform Drivers About Intentions of Their Automated Cars. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Ann Arbor, MI, USA) (*AutomotiveUI '16 Adjunct*). Association for Computing Machinery, New York, NY, USA, 57–62.

- <https://doi.org/10.1145/3004323.3004329>
- [59] Alan M MacEachren, Robert E Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. 2012. Visual semiotics & uncertainty visualization: An empirical study. *IEEE transactions on visualization and computer graphics* 18, 12 (2012), 2496–2505.
- [60] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.
- [61] Natasha Merat, A. Hamish Jamson, Frank C.H. Lai, Michael Daly, and Oliver M.J. Carsten. 2014. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour* 27 (2014), 274–282. <https://doi.org/10.1016/j.trf.2014.09.005>
- [62] Jean Mercat, Thomas Gilles, Nicole El Zoghby, Guillaume Sandou, Dominique Beauvois, and Guillermo Pita Gil. 2020. Multi-head attention for multi-modal joint vehicle motion forecasting. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, IEEE, New York, NY, USA, 9638–9644.
- [63] Bonnie M Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.
- [64] Tobias Müller, Mark Colley, Gülsemin Dogru, and Enrico Rukzio. 2022. AR4CAD: Creation and Exploration of a Taxonomy of Augmented Reality Visualization for Connected Automated Driving. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 177 (sep 2022), 27 pages. <https://doi.org/10.1145/3546712>
- [65] Lace Padilla, Matthew Kay, and Jessica Hullman. 2022. *Uncertainty visualization*. Wiley, Hoboken, NJ, USA, Chapter 22, 405–421.
- [66] Bastian Pflöging, Maurice Rang, and Nora Broy. 2016. Investigating User Needs for Non-Driving-Related Activities during Automated Driving. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (Rovaniemi, Finland) (MUM '16)*. Association for Computing Machinery, New York, NY, USA, 91–99. <https://doi.org/10.1145/3012709.3012735>
- [67] Amir Rasouli and John K Tsotsos. 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (2019), 900–918.
- [68] Stephan Reuter, Ba-Tuong Vo, Ba-Ngu Vo, and Klaus Dietmayer. 2014. The Labeled Multi-Bernoulli Filter. *IEEE Transactions on Signal Processing* 62, 12 (2014), 3246–3260. <https://doi.org/10.1109/TSP.2014.2323064>
- [69] Katarzyna Samson and Patrycja Kostyszyn. 2015. Effects of cognitive load on trusting behavior—an experiment using the trust game. *PloS one* 10, 5 (2015), e0127680.
- [70] Katarzyna Samson and Patrycja Kostyszyn. 2015. Effects of Cognitive Load on Trusting Behavior – An Experiment Using the Trust Game. *PLOS ONE* 10, 5 (05 2015), 1–10. <https://doi.org/10.1371/journal.pone.0127680>
- [71] Sarah Schmidt and B Färber. 2009. Pedestrians at the kerb—Recognising the action intentions of humans. *Transportation research part F: traffic psychology and behaviour* 12, 4 (2009), 300–310.
- [72] Tobias Schneider, Joana Hois, Alischa Rosenstein, Sabiha Ghellal, Dimitra Theofanou-Fülber, and Ansgar R.S. Gerlicher. 2021. ExplAIn Yourself! Transparency for Positive UX in Autonomous Driving. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, Article 161, 12 pages. <https://doi.org/10.1145/3411764.3446647>
- [73] Brandon Schoettle and Michael Sivak. 2014. *A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia*. Technical Report. University of Michigan, Ann Arbor, Transportation Research Institute.
- [74] Andreas Th Schulz and Rainer Stiefelhagen. 2015. Pedestrian intention recognition using latent-dynamic conditional random fields. In *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, IEEE, New York, NY, USA, 622–627.
- [75] Santokh Singh. 2015. *Critical reasons for crashes investigated in the national motor vehicle crash causation survey*. Technical Report. National Highway Traffic Safety Administration.
- [76] Vanessa Stange, Anne Goralzik, Susanne Ernst, Markus Steimle, Markus Maurer, and Mark Vollrath. 2022. Please stop now, automated vehicle!—Passengers aim to avoid risk experiences in interactions with a crossing vulnerable road user at an urban junction. *Transportation research part F: traffic psychology and behaviour* 87 (2022), 164–188.
- [77] Jan Strohbeck, Vasileios Belagiannis, Johannes Müller, Marcel Schreiber, Martin Herrmann, Daniel Wolf, and Michael Buchholz. 2020. Multiple Trajectory Prediction with Deep Temporal and Spatial Convolutional Neural Networks. In *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, New York, NY, USA.
- [78] Richard M Taylor. 2017. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness*. Routledge, Abingdon, UK, 111–128.
- [79] Henry Togwell, Mark McGill, Graham Wilson, Daniel Medeiros, and Stephen Anthony Brewster. 2022. In-CAR Gaming: Exploring the Use of AR Headsets to Leverage Passenger Travel Environments for Mixed Reality Gameplay. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 369, 7 pages. <https://doi.org/10.1145/3491101.3519741>
- [80] Robert Tscharn, Marc Erich Latoschik, Diana Löffler, and Jörn Hurtienne. 2017. “Stop over There”: Natural Gesture and Speech Interaction for Non-Critical Spontaneous Intervention in Autonomous Driving. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (Glasgow, UK) (ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 91–100. <https://doi.org/10.1145/3136755.3136787>

- [81] Unity Technologies. 2020. *Unity*. Unity Technologies.
- [82] Anne Marthe Van Der Bles, Sander van der Linden, Alexandra LJ Freeman, and David J Spiegelhalter. 2020. The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences* 117, 14 (2020), 7672–7683.
- [83] Hanneke Hooft van Huysduynen, Jacques Terken, Alexander Meschtscherjakov, Berry Eggen, and Manfred Tscheligi. 2017. Ambient Light and Its Influence on Driving Experience. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Oldenburg, Germany) (*AutomotiveUI '17*). Association for Computing Machinery, New York, NY, USA, 293–301. <https://doi.org/10.1145/3122986.3122992>
- [84] Tamara von Sawitzky, Philipp Wintersberger, Andreas Riener, and Joseph L. Gabbard. 2019. Increasing Trust in Fully Automated Driving: Route Indication on an Augmented Reality Head-up Display. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (Palermo, Italy) (*PerDis '19*). Association for Computing Machinery, New York, NY, USA, Article 6, 7 pages. <https://doi.org/10.1145/3321335.3324947>
- [85] Gesa Wiegand, Malin Eiband, Maximilian Haubelt, and Heinrich Hussmann. 2020. “I’d like an Explanation for That!” Exploring Reactions to Unexpected Autonomous Driving. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) (*MobileHCI '20*). Association for Computing Machinery, New York, NY, USA, Article 36, 11 pages. <https://doi.org/10.1145/3379503.3403554>
- [86] Marc Wilbrink, Anna Schieben, and Michael Oehl. 2020. Reflecting the Automated Vehicle’s Perception and Intention: Light-Based Interaction Approaches for on-Board HMI in Highly Automated Vehicles. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (Cagliari, Italy) (*IUI '20*). Association for Computing Machinery, New York, NY, USA, 105–107. <https://doi.org/10.1145/3379336.3381502>
- [87] Philipp Wintersberger, Anna-Katharina Frison, Andreas Riener, and Tamara von Sawitzky. 2019. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *PRESENCE: Virtual and Augmented Reality* 27, 1 (2019), 46–62. https://doi.org/10.1162/pres_a_00320 arXiv:https://doi.org/10.1162/pres_a_00320
- [88] Philipp Wintersberger, Hannah Nicklas, Thomas Martlbauer, Stephan Hammer, and Andreas Riener. 2020. Explainable Automation: Personalized and Adaptive UIs to Foster Trust and Understanding of Driving Automation Systems. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) (*AutomotiveUI '20*). Association for Computing Machinery, New York, NY, USA, 252–261. <https://doi.org/10.1145/3409120.3410659>
- [89] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). Association for Computing Machinery, New York, NY, USA, 143–146. <https://doi.org/10.1145/1978942.1978963>