Highlights

User Gesticulation Inside an Automated Vehicle with External Communication can Cause Confusion in Pedestrians and a Lower Willingness to Cross

Mark Colley,Bastian Wankmüller,Tim Mend,Thomas Väth,Enrico Rukzio,Jan Gugenheimer

- Development of functional prototype enabling external communication of (simulated) autonomous vehicles via a mounted display.
- Online study with N=59 participants.
- Results indicate high reliance on factors independent of the external communication when users of such vehicles perform potentially confusing actions despite clear introduction.
- Implications on System, User, and Societal level are discussed.

User Gesticulation Inside an Automated Vehicle with External Communication can Cause Confusion in Pedestrians and a Lower Willingness to Cross

Mark Colley[a,*], Bastian Wankmüller[a], Tim Mend[a], Thomas Väth[a], Enrico Rukzio[a] and Jan Gugenheimer[b]

[a]*Institute of Media Informatics, Ulm University, James-Franck-Ring 8, Ulm, 89081, Germany*
[b]*Télécom Paris - LTCI, Institut Polytechnique de Paris, Paris, France*

ARTICLE INFO

Keywords:
External communication
interaction design

ABSTRACT

Automated vehicles are expected to require some form of communication (e.g., via LED strip or display) with vulnerable road users such as pedestrians. However, the passenger inside the automated vehicle could perform gestures or motions which could potentially be interpreted by the pedestrian as contradictory to the outside communication of the car. To explore this conflict, we conducted an online experiment (N=59) with different message types (no message, intention, command), gestures (no gesture, wave, stop), and user positions (driver, co-driver) and measured the pedestrian's confidence in crossing. Our results show that certain combinations (e.g., car indicates cross while the user in the driver seat gestures stop) confused the pedestrian, resulting in significantly lower confidence to cross. We further show that designing intention-based external communication led to less confusion and a significantly higher intention to cross.

1. Introduction

In recent years, the external communication of automated vehicles (AVs) has become a novel research field. This research is targeted towards replacing or enhancing interpersonal communication used today between human drivers and vulnerable road users or towards improving traffic safety for people with disabilities (Colley, Walch, Gugenheimer, Askari and Rukzio (2020b)). Various areas within this field have been explored: external communication modalities such as auditory (Colley et al. (2020b)), visual (Mahadevan, Somanath and Sharlin (2018)), or tactile concepts (Mahadevan et al. (2018)) have been proposed, trust (Holländer, Wintersberger and Butz (2019)) and scalability challenges (Colley, Walch and Rukzio (2020c)), and (some) legal issues (Inners and Kun (2017)) have been highlighted. Until now, mostly crossing scenarios were evaluated (Colley, Mytilineos, Walch, Gugenheimer and Rukzio (2020a)). Nevertheless, other scenarios such as the merging of bicyclists in front of AVs (Hou, Mahadevan, Somanath, Sharlin and Oehlberg (2020)), communication between an AV and a human driver in a different vehicle (Rettenmaier, Pietsch, Schmidtler and Bengler (2019)), or walking past a highly automated truck occupying a sidewalk (Colley et al. (2020a)) were investigated.

One rationale to include such communication is the possible absence of human passengers (Ackermans, Dey, Ruijten, Cuijpers and Pfleging (2020)). However, this communication will also be needed with a passenger present as the passenger could be engaged in a variety of non-driving related tasks in an automated journey (Pfleging, Rang and Broy (2016)). These include eating, sleeping, watching movies, (video) games, and fitness, amongst others (Pfleging et al. (2016)). Already, novel concepts including the vehicle's motion are integrated into Virtual Reality (VR) (Hock, Benedikter, Gugenheimer and Rukzio (2017)). The presence of a passenger alone could lead, however, to the interpretation of the vehicle being driven manually. In scenarios in which AV - pedestrian communication is then needed, the behavior of potential passengers inside the AV has not yet been accounted for.

*Corresponding author

mark.colley@uni-ulm.de (M. Colley); bastian.wankmueller@uni-ulm.de (B. Wankmüller); tim.mend@uni-ulm.de (T. Mend); thomas.vaeth@uni-ulm.de (T. Väth); enrico.rukzio@uni-ulm.de (E. Rukzio); jan.gugenheimer@telecom-paris.fr (J. Gugenheimer)

https://m-colley.github.io/ (M. Colley)

ORCID(s): 0000-0001-5207-5029 (M. Colley); 0000-0001-8431-6771 (B. Wankmüller); 0000-0002-4645-0280 (T. Mend); 0000-0003-3460-2248 (T. Väth); 0000-0002-4213-2226 (E. Rukzio); 0000-0002-6466-3845 (J. Gugenheimer)

Therefore, depending on the task, the human user/passenger could display a variety of movements such as nodding with the head, waving to interact with potential Augmented Reality (AR) or VR applications, or vehicle controls (Pickering, Burnham and Richardson (2007)). Communicating with people in- and outside of the vehicle is also possible. Additionally, research already proposed to use (static) gestures to navigate AVs (Qian, Ju and Sirkin (2020)). Gestures used for these activities could be misinterpreted by pedestrians unfamiliar with AVs. This could result in conflicts with the message displayed on the external Human-Machine Interface (eHMI): greeting a person with a raised hand could be perceived as a warning not to cross the street or a swipe gesture could indicate that it is safe to cross while the person is actually interacting with an AR headset that could be indistinguishable from glasses as AR glasses will be decreasing in size (e.g., see concept video of Magic Leap (Leap, 2015, second 16))) or with vehicle controls (e.g., gesture control of VW (Volkswagen (2020)) or Mercedes-Benz (Mercedes-Benz (2020))). Such misunderstandings can result in disapproval of eHMIs or even dangerous situations for the pedestrian. Furthermore, in AVs (Taxonomy (2014)), passengers can take every available seat, thus, they may sit on the driver's seat. This complicates a pedestrian's assessment of the situation as this position resembles "power" over the vehicle today.

Therefore, we conducted an online video-based study evaluating the aspects *User Position*, *Behavior*, and *eHMI concept* with N=59 participants. We found that participants were confused with the user's action despite having a clear introduction to the scenario with the clarification that the vehicle is driving automated. We also found that the intention-based message "Stopping" lead to lower confusion.

Contribution Statement: The presented findings contribute to the knowledge in the communication design of AVs by revealing problems when potentially contradicting behavior of the human user is taken into account. Insights from a user study show that even with a clear introduction of the vehicle being automated, the user's behavior within the AV makes a significant impact on the pedestrian's willingness to cross, and their confusion. We also found, in accordance with previous research (Ackermans et al. (2020)), that an eHMI has a positive impact on their willingness to cross. We conclude by discussing the impact of these insights for future eHMI design and a discussion of potential measures to counter the found effect.

2. Related Work

This work contributes to the field of eHMIs, addresses mode confusion, and includes aspects of gesture usage in vehicles. Therefore, an overview of these research fields is presented.

2.1. External Communication of Automated Vehicles

One proposed solution to overcome traffic-related problems, which are currently resolved via gestures and eye-contact (Rasouli, Kotseruba and Tsotsos (2017)), is to add an eHMI. These concepts can be grouped based on their modality, message type, and communication location. The message type refers to the information conveyed by the AV and can be distinguished into eight classes: Instruction, Command, Advisory, Answer, Historical, Predictive, Question, and Affective (Colley and Rukzio (2020)). The communication location parameter defines whether the communication occurs on the vehicle, the personal device, or the infrastructure (Colley and Rukzio (2020)). Additionally, situation parameters such as communication relationship (one-to-one, one-to-many, many-to-one, and many-to-many), acoustic noise, or communication partner (e.g., pedestrian or cyclist) must be considered (Colley and Rukzio (2020)). Work on eHMI focused on children (Deb, Carruth, Fuad, Stanley and Frey (2020); Charisi, Habibovic, Andersson, Li and Evers (2017)), people with vision impairments (Colley et al. (2020b); Colley, Walch, Gugenheimer and Rukzio (2019a)), general pedestrians (Ackermans et al. (2020); Dey, Martens, Wang, Ros and Terken (2018); Löcken, Golling and Riener (2019)), and bicyclists (Hou et al. (2020)). In this novel field, there are still issues such as overtrust (Holländer et al. (2019)) or scalability (Colley et al. (2020c)) to be addressed. While many factors such as the individual perception of AVs or external appearance (Ackermans et al. (2020)) are considered, the user within the AV is not involved in current evaluation scenarios. To the best of our knowledge, we are the first to *systematically* include the position and behavior of potential users in AVs in research on eHMIs.

2.2. Mode Confusion

Mode confusion has been the topic of research in various areas such as aircraft domain (Spencer Jr (2000); Degani, Shafto and Kirlik (1996)), flight guidance systems (Joshi, Miller and Heimdahl (2003)), service robots (Lankenau (2001)), and driver interfaces (Lee, Ahn and Yang (2014)). Cummings and Ryan define mode confusion in the context of AVs as a discrepancy in how the human driver believes the vehicle to operate and the actual vehicle behavior (Cummings and Ryan (2014)).

Degani and Kirlik describe mode-switches (e.g., switching from automated to manual flying or when control is taken over by the human user during an automated journey in an AV) in the aircraft domain to be intuitive when it is represented to the users (Degani and Kirlik (1995)), which is (with fewer distinction) comparable to the 6 automation levels of the Society of Automotive Engineers (SAE) (Taxonomy (2014)). Degani et al. categorize mode confusion into two areas: (1) misidentification of automation behavior, meaning that the automation behaves differently to the expectations of the user, and (2) wrong assumption of automation mode when the user operates in the belief that a different mode is active (Degani et al. (1996)), the latter being in line with the definition of Norman (Norman (1983)). Leveson, Pinnel, Sandys, Koga and Reese (1997) distinguish six mode confusion problems: Interface Interpretation Error, Inconsistent Behavior, Indirect Mode Change, Operator Authority Limits, Unintended Side Effects, and Lack of Appropriate Feedback.

Kurpiers and Biebl (Kurpiers, Biebl, Mejia Hernandez and Raisch (2020)) distinguish between subjective and objective mode confusion measurements. While giving fast and explicit information on mode awareness, the authors state that subjective measures inherently have flaws such as personal bias, and misunderstandings of the system under evaluation can not be directly unveiled. Objective measurement methods include measuring gaze behavior, employment of non-driving task related tasks and measuring engagement in them, behavior patterns, and facial expressions. Nevertheless, the main interest in mode awareness lies in the consequential behavior adjustment (Kurpiers et al. (2020)).

Mode confusion is difficult to measure (Kurpiers et al. (2020)). Often, interviews are used (Johnson and Pritchett (1995); Kurpiers et al. (2020)). Other measurements include time to recognize a problematic mode (Johnson and Pritchett (1995)) or having participants determine which mode the system currently is in (Lee et al. (2014)).

Cognitive load (or lockup) was hypothesized to being associated with mode error (Spencer Jr (2000)). Due to the need to resolve divergent mode capability assumptions, cognitive resources are necessary. Therefore, we believe cognitive load to increase with mode confusion. Cunningham and Regan (Cunningham and Regan (2015)) primarily focus on the issue of overtrust of drivers in the AV. However, they show trust to be a contributing factor to mode confusion.

In the field of eHMIs, the main operator is the user within the AV. However, with pedestrians or cyclists communicating with the AV, these become exposed to potential mode confusion issues when the vehicle state is not clear. For the pedestrian, knowledge about which of the six automation levels (Taxonomy (2014)) is currently active is irrelevant, however, the binary information *vehicle drives automatedly* vs. *vehicle is driven manually* (see wrong assumption of automation mode (Degani et al. (1996))) is highly important.

2.3. Gesture Usage in Vehicle

Gestures can be divided into static and dynamic (Reifinger, Wallhoff, Ablassmeier, Poitschke and Rigoll (2007)). These differ in whether the arm and/or hand move. For manual driving, the premise is to keep the eyes on the road and the hands on the wheel (González, Wobbrock, Chau, Faulring and Myers (2007)). Nevertheless, gesture interfaces found their way into the vehicle. Riener, Ferscha, Bachmair, Hagmüller, Lemme, Muttenthaler, Pühringer, Rogner, Tappe and Weger (2013) investigated gestural interaction and showed that most interactions are performed in close proximity to the steering and shifting regions.

Qian et al. (2020) explored in-air static hand gestures for the “final 100 meter” problem of AVs. The “final 100 meter” problem stems from the potential need of AV users to adjust the direction of the Av to, for example, select the desired parking spot via speech or gesture due to, for example, uncertain user needs. They define six categories for user-defined hands shapes: Palm, fist, thumb, index finger, little finger, and roll. Most participants used a “palm-forward, fingers-up gesture – the same one that police use” (Qian et al., 2020, p. 6). Other gestures included waving. Participants in their study preferred dynamic gestures, however, the authors provide some evidence that static gestures could be better suited because of less performance time, less distraction, easier standardization, higher learnability, easier detection by technical systems, and less required performance space. While Qian et al. only investigated such control with their participants on the driver’s seat, this would (and should) also be possible when occupying other seats in AVs. The potential usage of such gestures could lead to confusion in conjunction with eHMIs.

3. Study

This study investigated the influence of gesture, eHMI, the passenger/user position, and their interactions on pedestrian behavior. The research question was as follows:

(a) First frame (passenger seat)
.

(b) 2nd frame: passenger, “cross” eHMI, wave.

(c) 3rd frame: driver, “stopping” eHMI, stop.

Figure 1: Three screenshots of various conditions used in the online study.

What impact do the variables *User Position*, *Behavior*, and *eHMI concept* have on (1) cognitive load, (2) trust, (3) attribution of control, (4) perceived safety, and (5) willingness to cross?

To study these influences, participants were presented with a video-based online study in which they encountered an AV with different eHMIs combined with varying user positions and gestures. Our hypotheses (H) regarding willingness to cross were:

H1: Pedestrians’ willingness to cross in front of a yielding AV increases when an eHMI is present (Ackermans et al. (2020); Matthews, Chowdhary and Kieson (2017); Habibovic, Lundgren, Andersson, Klingegård, Lagström, Sirkka, Fagerlönn, Edgren, Fredriksson, Krupenia et al. (2018)).
H2a: Pedestrians’ willingness to cross in front of a yielding AV will increase when the gesture of the user aligns with the message of the eHMI.
H2b: Pedestrians’ willingness to cross in front of a yielding AV will decrease when the gesture of the user contradicts the message of the eHMI.
H3: The effects on pedestrians’ willingness to cross in front of a yielding AV with the gesture alignment will be stronger when the user is sitting in the driver’s seat.

Regarding the other dependent variables, the study was exploratory.

In the study, each participant experienced **nine** conditions. This resembles a 3 × 3 design with the independent variables *Gesture Type* and *eHMI concept*. The *Gesture Type* consisted of the levels *none*, *wave*, and *stop*. The *eHMI concept* was split into the levels *intention “Stopping”* (23049:2018 (2018)), *command “Cross”* (Mahadevan et al. (2018); Colley et al. (2020b,a)), and *none*. Research is still not clear about the messages eHMIs should convey (Moore, Currano, Strack and Sirkin (2019); Colley and Rukzio (2020); Colley et al. (2020c,b); Dey, Habibovic, Löcken, Wintersberger, Pfleging, Riener, Martens and Terken (2020); 23049:2018 (2018); Mahadevan et al. (2018)), therefore, two of the most common approaches (Colley and Rukzio (2020)) were used. As we were interested in the effect of the position of the human user, we employed the factor *User Position* with the levels *driver’s seat* and *passenger’s seat* (or co-driver’s seat) as a between-subject factor.

The messages we present to the user were all indicating the ability to cross either through signaling an intention of the car (*intention “Stopping”*) or suggesting an action to the pedestrian (*command “Cross”*). While we were also interested in exploring an indicator against the ability to cross such as “Don’t Cross” or “Not stopping”, these message types are currently not intended to be used in future AVs and are unexplored in research. The main reason here is that there are few scenarios in which an AV would try to communicate with this negative framing. While the car itself could use this communication to warn pedestrians that it is driving past them (see “Crosswalk chicken” game between AVs and pedestrians as described by Ball (Millard-Ball (2018)), we believe it would mainly use these commands in mixed traffic scenarios to communicate a potential danger of a third vehicle. Due to the uncertainty of this negative framing we did not include these messages in our study design. Each session started with the introduction and the agreement of informed consent. The introduction to the study was as follows (including boldness):

> *You will see* ***nine*** *videos of a highly automated vehicle approaching you. Future automated vehicles could have an attached display, for example, on the grille to communicate with pedestrians. The vehicle* ***steers, brakes, and accelerates*** *(lateral and longitudinal guidance).*

> *Passengers will be able to engage in* ***every imaginable activity****. You are supposed to imagine standing at the curb and to assess whether you would be willing to cross in front of the highly automated vehicle.*

Imagining future (mixed) traffic is probably difficult for participants. Therefore, we clearly highlighted the capabilities and responsibilities of the AV. We also highlighted the possibility for the human user to engage in non-driving related tasks (Pfleging et al. (2016)). The human user carried a Jins Meme (Meme (2020)) as a surrogate for future AR headsets. The conditions were presented in randomized order. The eHMI was explained prior to every condition: “The vehicle will show "Cross"[/or "Stopping"] on the attached display when it wants to let you cross.” After each video, participants were asked to rate the measurements described in Section 3.2. A session lasted ≈ 25 min; participants were compensated with 3.7 *currency*.

3.1. Apparatus

(a) Ghost driver costume. (b) Front view of vehicle.

Figure 2: Apparatus for the video capture.

We developed a prototype (see Figure 2) which enabled us to display any message on a display mounted to the vehicle’s grille. The driver can be hidden via the *Ghost driver protocol* (Rothenbücher, Li, Sirkin, Mok and Ju (2016)) (see Figure 2a). Inside the vehicle, a Raspberry Pi 4 is connected to a touch display. This display enables the driver to define the timing of displaying an eHMI on the mounted display. This can be achieved manually or can be configured by defining a speed range at which the eHMI should be displayed. The speed of the vehicle is read via OBD-2. For the videos, we used the traffic practice area in Neu-Ulm, Germany. The vehicle drove approximately 30km/h in the videos before decelerating. A Canon EOS 650D was fixed on a tripod at the height of 1.80 m. We opted for videos taken in the real compared to a virtual world to display lighting conditions realistically. The display was later overlaid with text that resembled the real lighting conditions. This allowed us to use the same video for equal configuration when only the eHMI changed. Thus, the vehicle’s movement was exactly the same for each seating * gesture configuration. The same two authors drove the vehicle and trained beforehand to do so with little variance.

3.2. Measurements

A real-world study including objective measurements was not possible due to the potentially severe consequences in case of misinterpretation. Therefore, we operationalized mode confusion with the following dependent variables. Participants were asked to indicate their cognitive load using the mental workload subscale of the raw NASA-TLX (Hart and Staveland (1988)) 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). *Predictability/Understandability* (called *Understanding* from here) and *Trust* were measured using the *Trust in Automation* questionnaire by Körber (Körber (2019)). 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 the same 5-point Likert scales on two statements (“I trust the system.” and “I can rely on the system.”). Understanding and trust are expected to decrease when mode confusion occurs.

Additionally, participants rated their perceived safety (PS) using four 7-point semantic differentials (Faas, Kao and Baumann (2020a)). PS was already used to measure the effects of eHMIs (Faas, Mathis and Baumann (2020b)) and, in particular, the effect of displaying the driving mode (Joisten, Alexandi, Drews, Klassen, Petersohn, Pick, Schwindt and Abendroth (2020)), however, without including a human user inside the vehicle in automated mode. On 7-point Likert scales, we asked participants whom they believed to be in control ("Control"; *1=Person* to *7=Vehicle*) and whom their crossing decision was based on ("Decision Factor"; *1=Person* to *7=Display*). To directly measure potential mode confusion, we asked participants about their agreement to the statement (inspired by Johnson and Pritchett (1995)) *While the vehicle was communicating, the behavior of the human passenger confused me; "Confusion"* and whether the actions and intentions of the vehicle ("vehicle mode clarity") and the human user ("passenger mode clarity") were clear (each one question; *1=Totally Disagree* to *7=Totally Agree*). We also showed participants three screenshots (see Figure 1) of the video per condition: (1) at the start, (2) directly when the eHMI was starting to be visible, and (3) when the vehicle had stopped. For each screenshot, participants rated their willingness to cross (*1=completely unsure* to *7=completely sure*). We believe that willingness to cross declines with increased mode confusion. A single item assessing the driving style (*1=completely safe* to *7=completely dangerous*) was also included. Participants were also asked whether they would cross the street in this scenario *during* the approach and *after* the vehicle stopped (yes/no).

After all videos, we posed open questions regarding feedback and improvement proposals. Additionally, participants ranked the scenarios in terms of willingness to cross and confusion. Immersion was rated using the *Immersion* subscale of the Technology Usage Inventory (Kothgassner, Felnhofer, Hauk, Kastenhofer, Gomm and Kryspin-Exner (2013)) showing medium immersion (M=17.78, SD=5.53).

3.3. Participants

We recruited 72 participants via Prolific. These were pre-selected to being from the USA to avoid potential confounding variables such as left-hand vs. right-hand traffic or culture (Rasouli and Tsotsos (2019)). We had to exclude 13 participants due to failed attention checks resulting in N=59 participants (34 male, 25 female) aged 18-67 (M=33.54, SD=11.15). 32 participants remained in the *user position* driver group (13 female, 19 male, M=34.50, SD=13.12), 27 in the *user position* passenger group (12 female, 15 male, M=32.41, SD=8.34). A Mann-Whitney-U-Test showed no significant differences in the age (p=.99). A chi-square test of independence was performed to examine the relation between gender and the *user position*. Gender distribution also did not significantly differ ($\chi^2(1, N = 59) = 0.001$, p=.98). Participants showed medium *Propensity to Trust* (Körber (2019)) with the mean being M=2.89 (SD=.57). A Mann-Whitney-U-Test showed no significant differences in *Propensity to Trust* (p=.1) between *user positions*.

4. Results

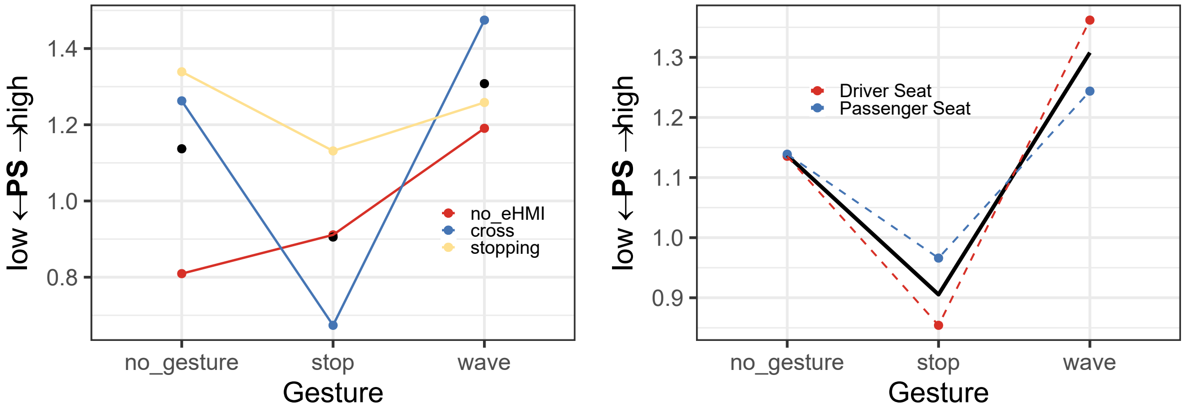
We focus on the main and interaction effects (in figures referenced as IE) of the three independent variables *user position* (between-subject), *gesture*, and *eHMI* (both within-subject). For non-parametric data, we used *nparLD* (Noguchi, Gel, Brunner and Konietschke (2012)), a method shown to be robust for unequal group sizes (Colley et al. (2020b)). ANOVA-type statistics are reported and Bonferroni correction was used for post-hoc tests. For main effects, we show the effect as a black line and the accompanying dashed lines mainly for the between-subjects factor *user position*. Effect sizes were calculated using Rosenthals's formula (Rosenthal, Cooper and Hedges (1994)).

4.1. Perceived Safety, Trust, & Cognitive Load

The non-parametric variance analysis (NPVA) revealed a significant main effect on PS of eHMI (F=4.08, df=1, p=.02) and gesture (F=6.48, df=1, p=.002, see Figure 3b). Pairwise comparisons using Dunn's test revealed no significant differences for the levels of eHMI but revealed a significant difference (p=.02, Z = -2.50, r=-0.13) between the gesture *stop* and *wave*. Participants felt less safe when the user displayed the *stop* gesture.

The NPVA also showed a significant interaction effect for *gesture* × *eHMI* (F=5.08, df=1, p=.001; Figure 3a). PS was mostly affected when the eHMI showed *cross* and the human user (falsely) was interpreted as showing a *stop* gesture. This indicates that the conveyed (interpreted) contradictory messages reduce perceived safety. This was not the case for the intention-based message "Stopping", which we assume is less prone to be seen as a contradictory message.

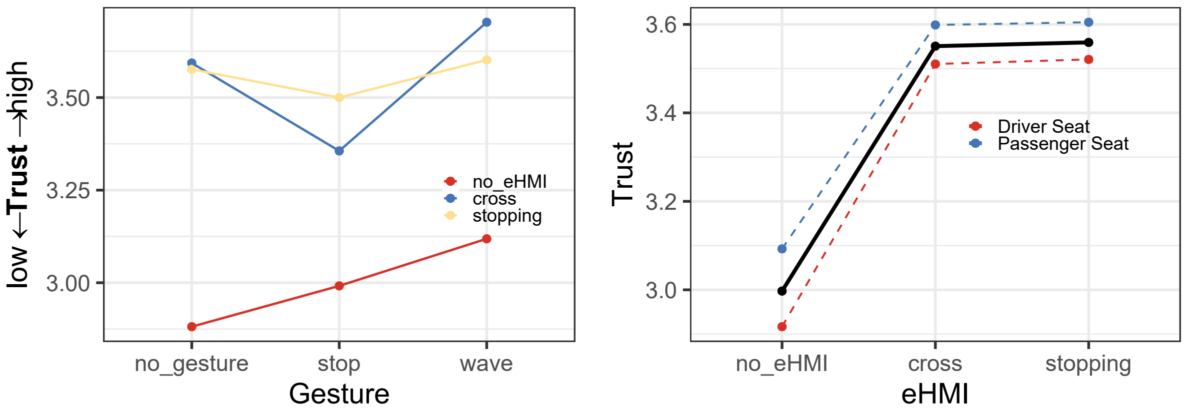
For the *Trust in Automation* questionnaire (Körber (2019)), the NPVA showed significant main effects of eHMI (F=34.24, df=1, p<.001; see Figure 4d) and gesture (F=8.97, df=2, p<.001; see Figure 4c) on the understanding subscale. Pairwise comparisons using Dunn's test revealed a significant difference between the eHMIs *no eHMI* and *cross* (p<.001, Z = 7.25, r=0.38) as well as between *no eHMI* and *stopping* (p<.001, Z = -7.54, r=-0.40). An eHMI



(a) IE of gesture × eHMI.

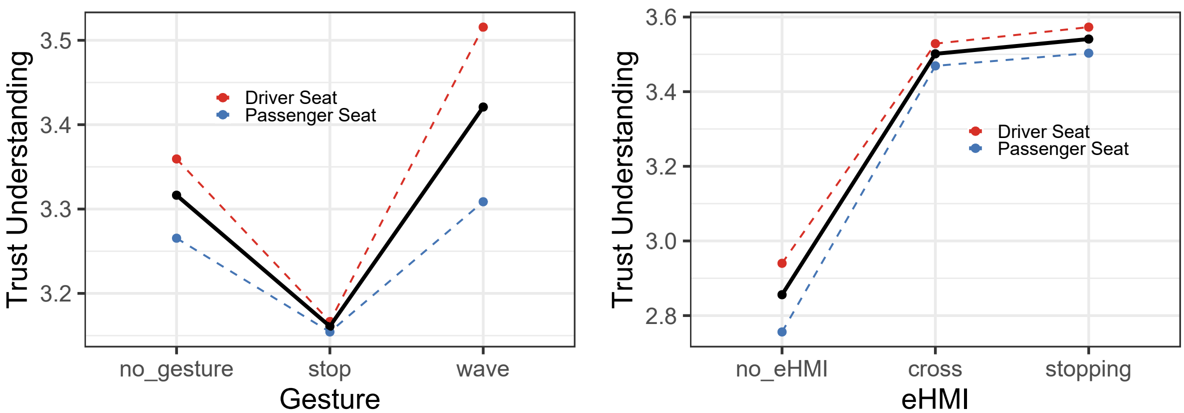
(b) Effect of gesture.

Figure 3: Effects on PS.



(a) IE of gesture × eHMI on trust.

(b) Effect of eHMI on trust.



(c) Effect of gesture on understanding.

(d) Effect of eHMI on understanding.

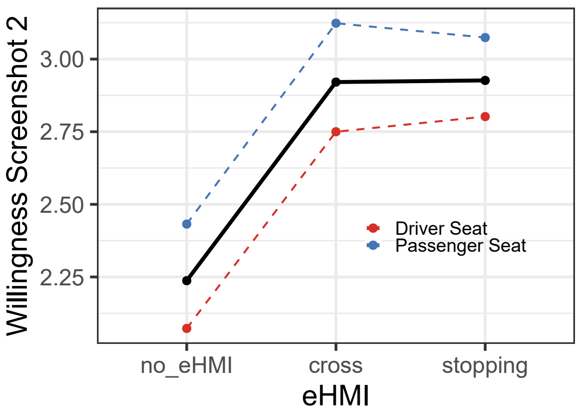
Figure 4: Effects on trust and understanding.

increased understanding significantly. For the gestures, Dunn's test revealed a significant difference (p=.0035) between *stop* and *wave* (Z = -3.04, r=-0.16). Waving increased subjective understanding. However, waving also reduced attribution of control towards the AV, showing potential confusion.

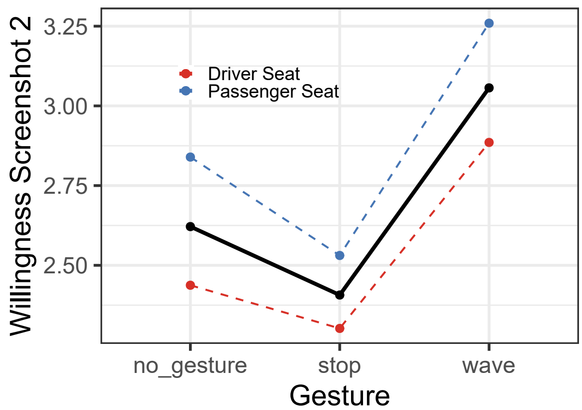
The NPVA also showed a main effect of eHMI (F=23.61, df=2, p<.001; see Figure 4b) and gesture (F=4.07, df=2, p=.02) on the trust subscale. Pairwise comparisons using Dunn's test again revealed significant differences between *no eHMI* and *cross* (p<.001, Z = 5.50, r=0.29) as well as between *no eHMI* and *stopping* (p<.001, Z = -5.50, r=-0.29). Trust was increased with an eHMI present. For the gestures, Dunn's test did not reveal a significant difference. Additionally, the NPAV found a significant interaction effect of gesture × eHMI (F=2.83, df=4, p=.02) on trust (see Figure 4a). Trust increased with a gesture when no eHMI was present. For the intention-based "Stopping" message, the Trust score remained approximately equal, only being slightly reduced with the stop gesture. For the command-based "Cross" eHMI, the combination with the stop gesture reduced Trust more. In line with the results of perceived safety (see Figure 3a), we assume that the clear contradiction between the two commands is responsible for this.

The NPVA also showed a significant main effect of gesture on mental load (F=5.43, df=2, p=.004). Pairwise comparisons using Dunn's test, however, revealed no significant effect between the levels of gestures.

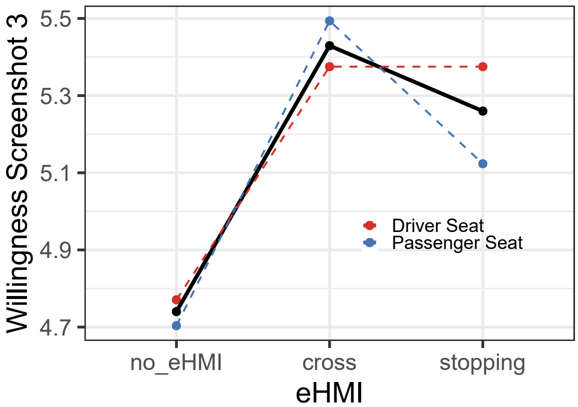
4.2. Willingness to Cross



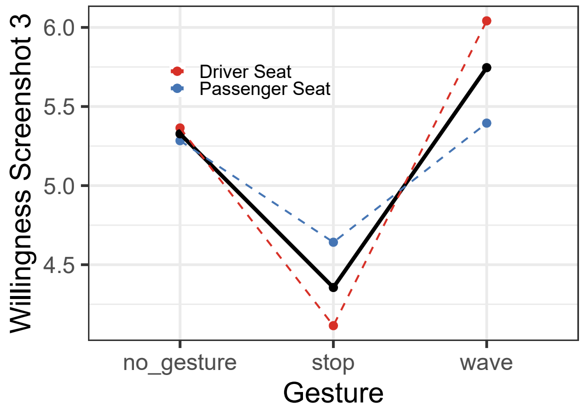
(a) Effect of eHMI on willingness in screenshot 2.



(b) Effect of gesture on willingness in screenshot 2.



(c) Effect of eHMI on willingness in screenshot 3.



(d) Main effect of gesture (black) and IE (dashed lines) of gesture × user position on willingness in screenshot 3.

Figure 5: Effects on willingness to cross in screenshots.

The NPVA found no significant differences in the willingness to cross for the first screenshot. However, the NPVA found significant main effects of eHMI and gesture for the second and the third screenshot (see Figure 5). Post-hoc comparisons using Dunn's test revealed, for the second screenshot, a significant difference between the eHMIs *no eHMI* to *cross* (p=.004, Z = 3.04, r=0.16) and to *stopping* (p=.001, Z = -3.36, r=-0.18). Additionally, a significant difference was found between the gestures *stop* and *wave* (p=.01, Z = -2.72, r=-0.14). Participants indicated higher willingness to cross in the second screenshot when an eHMI was present and when no *stop* but a *wave* gesture was used. This is reasonable as the pedestrian does not have to rely on the assumed behavior but can rely on the communication of the eHMI. Also, the gesture wave supports the decision to cross while stop is a counter-argument.

For the third screenshot, Dunn's test revealed the same differences for the eHMI (see Figure 5c; *no eHMI* - *cross* (p=.002, Z = 3.29, r=0.17), *no eHMI* - *stopping* (p=.01, Z = -2.67, r=-0.14)). For gesture, it revealed a significant difference between all levels (*no gesture* - *stop* (p<.001, Z = 3.78, r=0.20); *no gesture* - *wave* (p=.01, Z = -2.61, r=-0.14); *wave* - *stop* (p<.001, Z = -6.39, r=-0.34)). Willingness to cross was highest with the user waving, medium with no, and lowest with the stop gesture.

Additionally, an interaction effect of *gesture* × *user position* was shown (F=3.34, df=2, p=.04; Figure 5d). The plot shows that the *stop* gesture has a smaller effect on the willingness to cross in screenshot 3 when the human user is on the passenger's side. This indicates that gestures performed on the driver's seat are more relied upon and more power is attributed to the user there. This is sensible for countries with right-hand traffic. No other main or interaction effects were found for the willingness to cross.

4.3. Mode Confusion

Regarding (mode) confusion measurements, the NPVA found a significant effect of gesture on confusion (F=29.07, df=2, p<.001; see Figure 6a). Dunn's test showed a significant difference in mode confusion between *no gesture* and *stop* (p<.001, Z = -7.71, r=-0.41) and between *stop* and *wave* (p<.001, Z = 7.88, r=0.42). Mode confusion was high when the user employed the *stop* gesture. Again, this correlates to the effects on Trust and perceived safety and is attributed to the contrary message statements of the gesture and eHMI.

The NPVA found a significant interaction effect of *gesture* × *eHMI* on confusion (F=2.72, df=3, p=.037; Figure 6d).

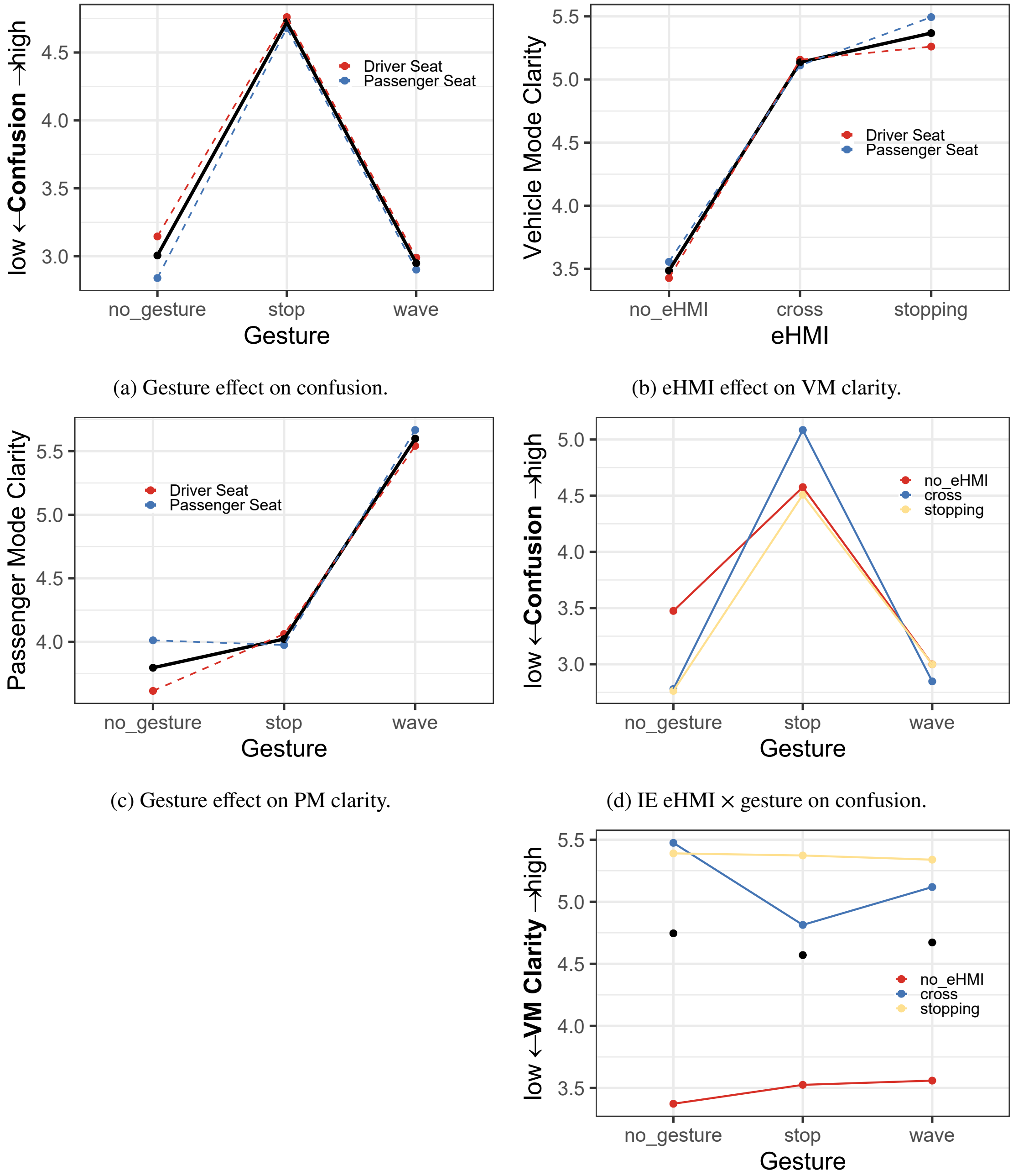
The NPVA also revealed a significant effect of eHMI (F=47.09, df=2, p<.001; see Figure 6b) and the interaction of *gesture* × *eHMI* (F=33.90, df=4, p=.04; see Figure 6e) on vehicle mode clarity. Dunn's test revealed the significant difference between *no eHMI* and *cross* (p<.001, Z = 8.01, r=0.43) as well as *no eHMI* and *stopping* (p<.001, Z = -9.11, r=-0.48). Presence of a eHMI significantly increased vehicle mode clarity (no eHMI: M=3.50, SD=1.88; cross: M=5.14, SD=1.63; stopping: M=5.37, SD=1.43).

Furthermore, it found a significant effect of gesture (F=33.90, df=2, p<.001; see Figure 6c) as well as a significant interaction effect of *eHMI* × *user position* (F=4.52, df=2, p=.01; see Figure 7) on passenger mode clarity. Dunn's test showed significant differences between the gestures *no gesture* and *wave* (p<.001, Z = -8.90, r=-0.47) and *stop* and *wave* (p<.001, Z = -7.68, r=-0.41). Subjective passenger mode clarity was higher when the user waved.

Both interaction effects of *eHMI* × *gesture* (see Figure 6d and e) show that displaying the command *cross* on the eHMI in conjunction with the misleading gesture *stop* leads to the highest confusion and a drop in assessed vehicle mode clarity. Figure 7 shows an inverse relationship between user position and eHMI on passenger mode clarity when *cross* is displayed. While perceived clarity is high when the human user is on the passenger's seat, it is low when the user is on the driver's seat. We believe this to be a result of current beliefs. Today, the driver is in full control. Although being introduced as automated, the notion of having control when sitting on the driver's seat still seems to be strong.

4.4. Attribution of Control, Driving Style, and Decision Factor

Despite being introduced as a highly automated vehicle taking over lateral and longitudinal guidance, the NPVA found a significant effect of user position (F=4.70, df=1, p=.03), eHMI (F=26.26, df=2, p<.001), and gesture (F=13.18, df=2, p<.001) on control attribution. Multiple comparisons using Dunn's test revealed a significant difference between user position *passenger* and *driver* (p<.001, Z = -4.62, r=-0.20). Control was significantly more attributed to the vehicle when the human user was on the passenger's side. This is sensible and shows that despite the introduction as automated, the current assumptions about vehicles matter more than a simple introduction. This highlights the necessity for a sensible information campaign for the general public once AVs become reality. Dunn's test also showed a significant difference between *no eHMI* and *cross* (p<.001, Z = 4.47, r=0.24) and between *no eHMI* and *stopping* (p<.001, Z = -4.30, r=-0.23). The presence of an eHMI led to the attribution of control to the vehicle. Dunn's test also revealed a significant difference in control between *no gesture* and *stop* (p=.003, Z = 3.10, r=0.16)



(a) Gesture effect on confusion.

(b) eHMI effect on VM clarity.

(c) Gesture effect on PM clarity.

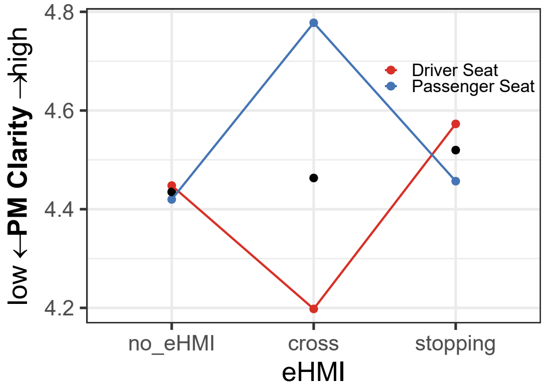
(d) IE eHMI × gesture on confusion.

(e) IE eHMI × gesture on VM clarity.

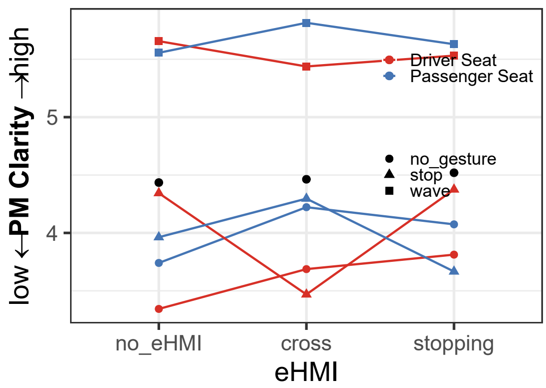
Figure 6: Main and IE on mode confusion scales (confusion and **vehicle** mode (VM) clarity).

and between *no gesture* and *wave* (p<.001, Z = 3.70, r=0.20). A gesture lead to the lower attribution of control to the vehicle.

The NPVA also found a significant interaction effect of *eHMI* × *gesture* (F=3.95, df=4, p<.01; Figure 8). While the introduction clearly stated that the vehicle is in control and the user can *be engaged in every activity imaginable*, these gestures still significantly decreased attributed control in the vehicle.



(a) IE eHMI × position on PM clarity.



(b) IE eHMI × position on PM clarity per gesture.

Figure 7: IE on **passenger** mode (PM) clarity).

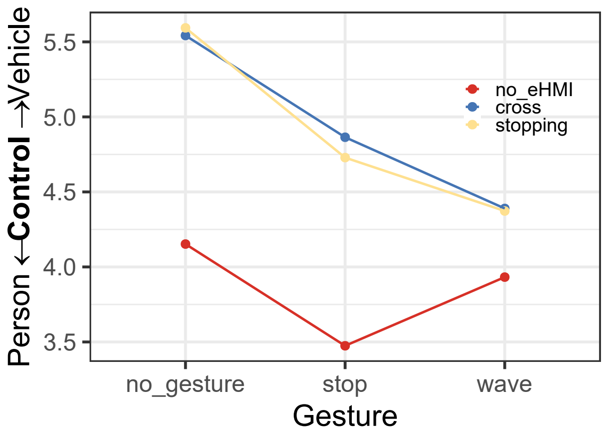


Figure 8: IE of eHMI × gesture on control.

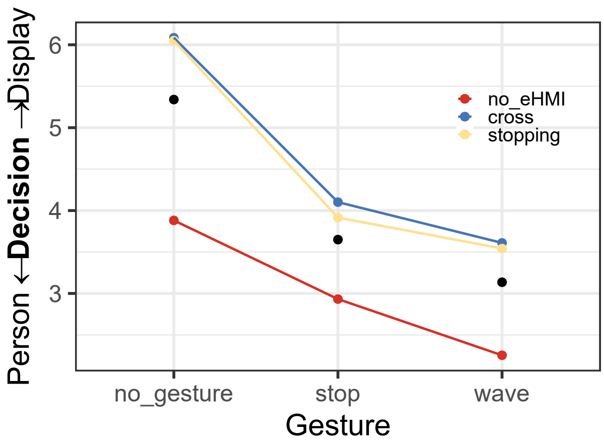
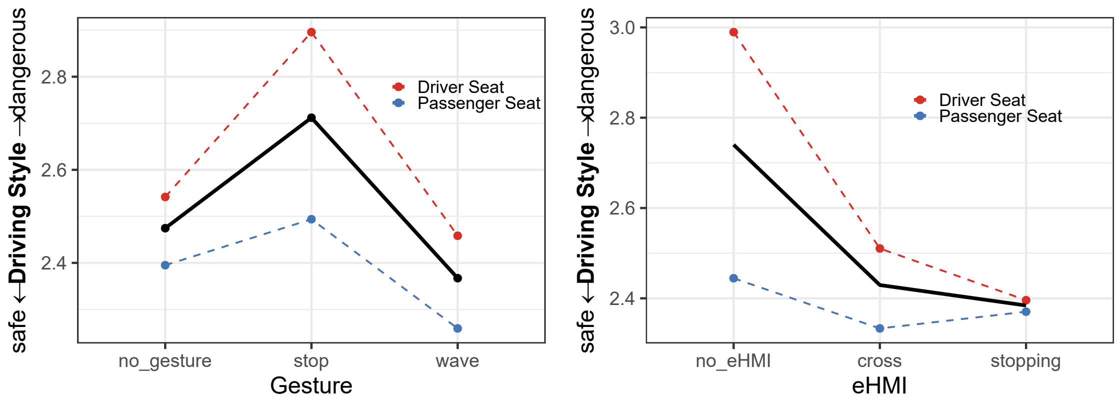


Figure 9: IE of eHMI × gesture on decision.

The NPVA showed a highly significant main effect of user position (F=7.46, df=1, p=.006), eHMI (F=40.21, df=2, p<.001), and gesture (F=48.90, df=2, p<.001) on the decision factors to cross. Dunn's test showed significant differences between *no eHMI* and *cross* (p<.001, Z = 6.25, r=0.33) and *stopping* (p<.001, Z = -5.86, r=-0.31). The decision was highly based on the presence of the eHMI. Dunn's test also showed a significant difference between *no gesture* and *stop* (p<.001, Z = 6.76, r=0.36) and *cross* (p<.001, Z = 8.80, r=0.47). Differences between *stop* and *wave* almost reached significance (p=.06, Z = 2.04, r=0.11). With no gesture, participants reported to rely on the display. With the gesture, more emphasis was laid on the gesture, the *wave* gesture being reported to be most influential. Regarding user position, Dunn's test revealed a significant difference (p<.001, Z = 6.76, r=0.29) between *passenger* and *driver*. Sitting on the passenger's seat lead to basing one's decision more on the display.

The NPVA also showed a significant interaction effect of *gesture* × *eHMI* on the decision factors to cross (F=2.99, df=4, p=.02; see Figure 9). With no gesture and an eHMI present, participants clearly based their decision on the display/eHMI. With the introduction of gesture, however, they became indecisive (no gesture with cross eHMI: M=6.08, SD=1.59; wave gesture with cross eHMI: M=3.61, SD=2.10; see Figure 9). This shows that more emphasis was still put on the eHMI, but that participants seemed to be confused and took both the eHMI and the passenger('s actions) into account.



(a) Effect of **gesture**.

(b) Effect of **eHMI**.

Figure 10: Main effects on driving style assessment.

The NPVA showed highly significant main effect of gesture (F=3.35, df=2, p=.036; Figure 10a) and eHMI (F=4.59, df=2, p=.01; Figure 10b) on the driving style assessment. Dunn's test found no significant differences for the eHMI levels, however, found a significant difference between the gestures *stop* and *wave* (p=.02, Z = 2.47, r=0.13). Showing the stop gesture lead to the impression of a significantly more dangerous driving style.

4.5. Decision to Cross

As crossing relies on a decision which is ultimately a boolean (yes/no) value, we included a yes/no question on whether the participant would cross during the approach and after the vehicle stopped. For the categorical data of whether the participant would have crossed *during* the approach or *after* the vehicle stopped, we employed a generalized linear mixed model (GLMM). As every participant rated multiple AV approaches, data was of hierarchical nature (measurements nested within participants). Therefore, we specified the participant as a random effect to account for individual differences in our model. Predictors in each model were the following: user position (dummy coded with the *driver seat* being the reference category), gesture (dummy coded for *no gesture* being the reference category), eHMI (dummy coded for *no eHMI* being the reference category) as well as all interaction terms. However, this was only possible for the *approach* data as the singularity assumption was violated for the *stopped* data (θ = 3.24) (author (2020)). Therefore, only descriptive values are reported for this (see Table 1).

According to Figure 11, the gesture *wave* and the eHMIs *stopping* and *cross* have a significant positive impact on the decision to cross during the approach. The largest significant effect against crossing during the approach is shown for the interaction between the eHMI *cross* and the gesture *stop*. Nevertheless, the gesture *wave* in combination with the eHMIs *stopping* or *cross* also had a significant negative effect on the crossing decision. We interpret that a

	During Approach		After Stopped	
Variable number of answers	Yes	No	Yes	No
eHMI				
no eHMI	40 (22.60%)	137	156 (88.14%)	21
Cross	56 (31.64%)	121	161 (90.96%)	16
Stopping	63 (35.59%)	114	158 (89.27%)	19
Gesture				
no gesture	49 (27.68%)	128	165 (93.22%)	12
wave	71 (40.11%)	106	174 (98.31%)	3
stop	39 (22.03%)	138	136 (76.84%)	41
User Position				
Driver seat	85 (29.51%)	203	256 (88.89%)	32
Passenger seat	74 (30.45%)	169	219 (90.12%)	24

Table 1
Participants' decision of whether to cross in relation to the factors both for during the approach of the vehicle and after the vehicle stopped. Values differ for the *User Position* due to different number of participants.

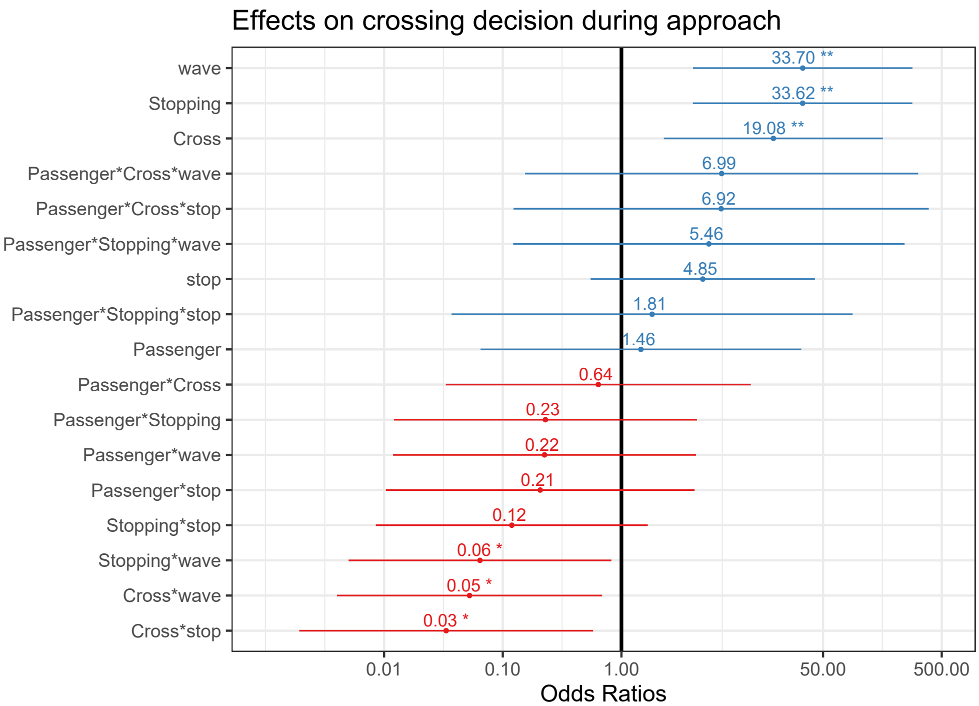


Figure 11: GLMM for the crossing decision during the approach (gestures with lower case letters). The odds ratio defines the probability that a participant would have crossed during the approach of the AV. The odds ratio of 33.70 for the gesture *wave* shows that the probability of crossing during the approach was 33.70 times higher with the wave gesture than without. On the other end, the probability to cross during the approach was only 0.03 times as probable when the eHMI *Cross* and the *stop* gesture were employed.

gesture per se with an eHMI resulted in fewer actual positive decisions to cross during the approach. This is interesting as results showed a significant positive impact on willingness to cross during the approach (i.e., screenshot 2) when the gesture *wave* was used and when an eHMI was active (see Figure 5a and b). Willingness to cross and the actual decision, therefore, seem not necessarily to correlate. This is relevant as previous research employed sliders to measure willingness to cross without including the actual decision as dependent variable (Dey, Walker, Martens and Terken (2019); Walker, Dey, Martens, Pfleging, Eggen and Terken (2019)).

4.6. Open Feedback

Participants were able to mention concerns and to provide improvement proposals at the end of the experiment. Five participants highlighted the low brightness of the display. We reflect on this in the Section 6.5. Two participants reflected on the intention-based communication. They found that "stopping" should become "stopped" after coming to a halt (P22, P51). One person indicated to have understood the user's movements inside the vehicle as necessary to the communication of the AV. P38 stated: "I do like the idea of people having to signal inside the car even if the car is in control." Others showed their confusion of the user behavior. P13 stated: "The "stop" hand gesture by the driver is very confusing. He appears to be ordering the pedestrian to stop, which causes the pedestrian to not cross the street."

5. Evaluation of Hypotheses

With this data, we were able to validate our hypotheses as follows:

- *H1: Pedestrians' willingness to cross in front of a yielding AV increases when an eHMI is present.*
  In line with previous work (Ackermans et al. (2020)), our results showed that the presence of an eHMI significantly improved participants' willingness to cross the street. Therefore, we accept our hypothesis.

- *H2a: Pedestrians' willingness to cross in front of a yielding AV will increase when the gesture of the passenger aligns with the message of the eHMI.*
  With regards to Figure 5 b and d, the willingness to cross for screenshot 2 and 3 were higher with the *wave*-gesture than compared to no gesture or even the contradicting *stop*-gesture. For the decision to cross during the approach of the vehicle, we also found a significant positive effect (see Figure 11) for the *wave*-gesture. However, there was a significant negative effect of eHMI cross* wave and stopping*wave on the decision to cross during the approach (see Figure 11). Therefore, we reject the hypothesis.

- *H2b: Pedestrians' willingness to cross in front of a yielding AV will decrease when the gesture of the passenger contradicts the message of the eHMI.*
  Figure 5 shows a negative effect of the *stop*-gesture on the willingness to cross. Figure 11 shows that the interaction between the eHMI showing "cross" and the gesture *stop* has a significant negative effect on the final decision to cross. However, this is not evident for the eHMI showing "stopping" in combination with the *stop* gesture. Therefore, we only partially accept this hypothesis.

- *H3: The effects on pedestrians' willingness to cross in front of a yielding AV with the gesture alignment will be stronger when the passenger is sitting in the driver's seat.*
  As shown in Figure 5d, the user position *driver* resulted in higher deflections with the gestures. However, we could not find a significant main effect. We also found no significant effect on the decision to cross during the approach (see Figure 11) that involves user position. Therefore, we partly accept our hypothesis with the constraint that this willingness is also dependent on the gesture.

6. Discussion

This study investigated the effect of *user position* in an AV, *eHMI* consisting of two textual representations and *user gestures*. The online study results showed that, despite the explicit introduction of the automated capabilities and the possibilities for a human user inside an AV, participants were confused when the human user inside the AV gesticulated. It also showed, on various dependent variables, the effects of contradictory (stop) and supporting (wave) gestures, which are irrelevant for the AV — pedestrian communication.

6.1. Mode Confusion

As mode confusion is difficult to measure (Kurpiers et al. (2020)), we employed several dependent measurements hypothesized to correlate with mode confusion. These were perceived safety, cognitive load, trust, understanding, willingness to cross, statements directly asking for confusion, attribution of control, experienced driving style, crossing decision during approach and after stopping, and the factor the crossing decision was based on. We interpret the results as clearly showing mode confusion when the human user displays potentially traffic-related gestures despite having introduced the vehicle as automated.

Participants' ratings showed high confusion when exposed to the stop gesture (see Figure 6a). Also, while ratings indicate high subjective passenger mode clarity with the wave gesture (see Figure 6c, M=5.60, SD=1.55), participants reported lower attributed control to the vehicle when this gesture was employed (M=4.23, SD=2.27 compared to M=5.10, SD=2.19 with no gesture) and based their decision less on the display than on the person (M=3.14, SD=2.19 compared to M=5.34, SD=1.92 without the gesture). Therefore, we interpret this as a clear mode confusion: while believing to understand the passenger and the movements, actually, these are misunderstood and the decision is consequently based on the wrong factors.

We could also show that the *user position* combined with a *gesture* has a significant impact on the willingness to cross (see Figure 5d). This should, however, not be the case. Again, we interpret this as showing mode confusion.

Finally, the GLMM (see Figure 11) showed a significant negative impact of the combination of the eHMI Cross with the gesture wave on the final decision to cross during the approach of the AV. However, both factor levels were shown to significantly heighten willingness to cross in isolation (see Figure 5a and b). We explain this by the unresolved confusion when the behavior and the eHMI are shown simultaneously.

6.2. Intention vs. Command Communication

Colley and Rukzio (2020) report that most publications on eHMIs either use an intention- or command-based message. The current technical report (23049:2018 (2018)) states that intention-based communication should be preferred. Our data supports this as we find no evidence that providing the "Cross" message results in any less confusion (see Figure 6a; Figure 11). *Stopping* does, however, lead to lower confusion (see Figure 6a), higher vehicle mode clarity (see Figure 6b), more consistent trust (see Figure 4a) and perceived safety (see Figure 3a) and seems to have no negative impact on control attribution (see Figure 8) as well as the decision factor (see Figure 9).

We used text as a communication modality as it was reported to be most unambiguous (Chang, Toda, Igarashi, Miyata and Kobayashi (2018); Deb et al. (2020)). Using other modalities such as animations or LED strips could reveal other conclusions; however, we believe that reliance on gesture behavior will be even stronger when using more ambiguous communication.

6.3. Gesture Usage in Automated Vehicles

Qian et al. (2020) propose to use static gestures to navigate an AV the last few meters (addressing the "Last 100 m Problem") to the final goal. Also, manufacturers already include gestures to interact with their vehicles. Additionally, with the advance of other activities within the AV (games, leisure, work), a wide range of gestures and, in general, user movement seems likely. Our work is one of the first to point towards issues arising with this development. While real-world data on how people will use AVs is missing and the prevalence of such gestures can not be estimated exactly, several consequences can be imagined. Mis- and even abuse can not be ruled out. Already today, Tesla users display a wide range of activities when in Autopilot mode (Edition (2018)). The user put his feet out of the window (Edition, 2018, second 3), some played games involving plastic lightsabers (Edition, 2018, second 78), slept (Edition, 2018, second 86)), and clapped (Edition, 2018, second 89). Our work draws from these developments. The experiment showed the confusion and mistrust towards the AV in conjunction with confusing user behavior. We showed that, despite a clear introduction to the AV's capabilities, pedestrian will, at least in early stages, take user behavior into account for their crossing decision. Such contradictions could, in turn, lead to a lack of acceptance and, thus, diminishment of the advantages AVs and their eHMIs. In extreme cases, the consequences could even be fatal.

6.4. Need for External Communication of Automated Vehicles

While there is still debate about whether (Moore et al. (2019)) and when (Colley et al. (2020a)) eHMIs are needed, our experiment shows that an eHMI helps the pedestrian in identifying who is in control of the driving task, provides a means on which pedestrians based their decision to cross on, supports their willingness to cross, and increases trust and their perceived safety. Especially in situations in which the pedestrian thinks the human user gesticulates not to cross, a deadlock can occur. This could lead to long delays and even traffic jams. Additionally, if it were to be unsafe to cross (e.g., because there is an approaching ambulance), waving gestures could be interpreted as "safe to cross", therefore, causing potential accidents. Table 1 also shows the effect of an eHMI during the approach of the AV: while only 22.60% would cross without an eHMI, this percentage increased to 35.59% with the eHMI "Stopping". However, this also shows that other factors, presumably most importantly implicit factors such as speed, are highly relevant for the crossing decision.

6.5. Limitations

The sample of N=59 participants was recruited from the USA. Therefore, transferability to other countries or cultures is not clear as culture plays an essential role in crossing decisions (Rasouli and Tsotsos (2019)). Additionally, our experiment was based on an online video-based study and lacked several real-world crossing characteristics. Video-based studies lack embodiment which was found to be a relevant factor for perceived presence in VR (Rogers, Funke, Frommel, Stamm and Weber (2019)) and was, therefore, proposed to be used in studying AV - pedestrian interaction (Colley, Walch and Rukzio (2019b)). Despite this limitation, a realistic prototype was developed and used for the videos which we believe to alleviate some of the drawbacks due to the high realism.

For our prototype, we used an LCD monitor with a contrast ratio of 1000:1, a viewing ratio of 178º, and a brightness of 250 cd/m^2 and used maximum brightness. However, as seen in Figure 1, this resulted in subperfect visibility. Therefore, we improved visibility via Adobe After Effects. However, we accounted for the requirements defined by Rettenmaier, Schulze and Bengler (2020). This indicates that for future developments, higher brightness and maximum contrast is necessary. Technical requirements for eHMIs are mostly unavailable, potentially because much work until now was done in Virtual Reality or was monitor-based (Colley et al. (2020c)).

To avoid confounding factors, we used as few videos as possible and altered the display's appearance via motion trackers and Adobe After Effects. Therefore, a single vehicle approached from the same side at approximately the same speed. This is a very constrained scenario which is lacking numerous factors of real-world scenarios (see scalability (Colley et al. (2020c))). Nonetheless, as we required different seating arrangements, there were some variations in speed and sound landscape.

Imagining a future with AVs was probably difficult for participants. Therefore, participants could have still been in the mindset of current traffic. However, it is unlikely that everybody closely follows the status of this technology. Without a broad awareness campaign, it seems likely that many people will be surprised in their first encounters with AVs. Thus, we argue that this scenario still depicts a realistic possibility.

Measuring mode confusion is difficult (Kurpiers et al. (2020)). Kurpiers et al. (2020) discourage purely subjective ratings for mode confusion measurements in AVs. However, in the presented study, objective measurements were not possible. Letting participants pass in front of the (simulated) AV could have fatal consequences, therefore, we opted for an online video-based study. However, this experiment should be repeated, for example, in VR to measure gaze and behavior patterns. Additionally, numerous measurements were performed using 7-point Likert scales, which potentially could have influenced the results. This was, for example for the attribution of control, done to allow for uncertainty to be communicated by the participant. However, a 7-point Likert scale could have made consistent answers more difficult.

7. Implications

The introduction of AVs will most likely have an enormous impact on traffic (Fagnant and Kockelman (2015)). As it is unlikely that every person knows about the state of the introduction of AVs and has enough knowledge about the different models, scenarios seem likely where the person encounters an AV without immediately knowing that it is an AV. Our experiment shows that *even when introducing the vehicle as automated*, participants were unsure about its state and how to interpret the behavior of the human user. While this might seem like an edge case, the consequences could be fatal. Therefore, it is necessary to address these. There are three levels on which this issue can (and should) be addressed: On the system, the user (or operator), and on the environmental/societal level.

7.1. System Level

By adjusting the AV to prevent or minimize the confusion potential, other counter-measurements might not be as indispensable. Simply not employing an eHMI or only using it in very specific instances(Moore et al. (2019)) is not supported by our data as the willingness to cross was positively affected by the presence of an eHMI. Norman (Norman (1983)) provided three ways to minimize mode confusion: (1) avoid modes, (2) mark modes distinctively, and (3) make commands per mode unique. As it is unlikely to avoid modes (i.e., different automation levels) as manufacturers such as Tesla incrementally improve their automated features (TeslaFi (2020)). Making commands unique is also no possibility regarding AV - pedestrian communication as the pedestrian does not command the AV. Therefore, marking modes distinctively seems like a reasonable approach. While it can be argued that having an external display could be a sufficient indication for the mode of the vehicle, the reported experiment invalidates this assumption. Ackermans et al. already investigated the effect of a conspicuous sensor on the effectiveness of eHMIs. In their discussion, they formulate "that an AV may need to visually call attention to its automated driving capability" (Ackermans et al., 2020, p. 10). As these sensors are being miniaturized (Chai, Nie and Becker (2021)), the question arises whether an additional signal is necessary. This is also important as companies such as Tesla do not incorporate such sensors in their design (Hecht (2018)). While, with high market penetration, AVs will become ubiquitous and pedestrians will likely be accustomed to them, such high penetration is not predicted to be achieved soon (LLP (2020)). Faas et al. (2020b) compared the effects of displaying status+ intent vs. only displaying intent vs. no eHMI. They conclude that "In summary, informing about the automated status improves trust, acceptance and UX [user experience]" (Faas et al., 2020b, p. 8). Our data support the need to include the visualization of the automation status.

Another attempt to approach the problem of pedestrian confusion is to make the human user (temporarily) invisible. One idea could be to employ windshields acting as one-way mirrors. Like this, the user within the vehicle can still look outside while the pedestrian does not see the human user inside. Variations of this attempt could be also be more promising. Instead of totally hiding the interior windows that become semi-transparent when unusual behavior of the human user is detected could have a smaller effect on pedestrian confusion. Another approach is to use the windshield as a display to communicate the intent of the vehicle, therefore, simultaneously blocking out human user gestures and actions (LUMINQ (2020)).

7.2. User and Pedestrian Level

The users of AVs with an attached eHMI have to be educated and sensitized about the impact of their movements and gestures. There are several ways to go on about the shown problem. Prohibiting certain activities seems to be unfavorable as this would diminish the advantages of AVs. The user could, however, be included in the communication process. For this, the AV could inform the user when potential confusion arises and stop the outward communication. It would then be the task of the human to communicate. Blocking out the user as described above seems to be favorable as human failure could be avoided. This blocking out, however, has to be accepted by the user.

For the pedestrian, the introduction of AVs with eHMIs requires some knowledge about the system's capabilities and requirements. The public must be educated about the advancement in AV technology and the meaning of such eHMIs (also discussed by Faas et al. (2020b)). As shown in this experiment, this is not a trivial task as, even with a clear introduction, participants were confused about the automation status. This could be made easier by making the human user invisible, as this is not a signal that has to be learned. Some research suggests that the adaptation process will be quick (Millard-Ball (2018)) and that pedestrians will even take advantage of AVs. Nonetheless, traffic will change and pedestrians will have to adjust their behavior. The driving style of an AV is still unclear or is subject to the user's liking (Yusof, Karjanto, Terken, Delbressine, Hassan and Rauterberg (2016)), especially when no person is inside. This could pose problems, especially on non-tech savvy people.

7.3. Environmental and Societal Level

Looking at the bigger picture of traffic, several implications could be drawn. The goal should be to minimize interactions between AVs and pedestrians. This can be achieved via alteration of the environment, for example, defining streets that AVs are allowed to use. Governments could explicitly regulate AVs routes. While this could pose inconveniences as the last few meters can not be driven automatedly, the interaction between AVs and pedestrians could be minimized. This could also strengthen local public transportation, reduce air pollution in cities, and, in general, reduce increased traffic as a possible rebound effect of the efficiency of AVs (Pakusch, Stevens, Boden and Bossauer (2018)). Additionally, streets explicitly for AVs could be marked or separated where possible. Again, this reduces the advantages of AVs and could be costly. Another approach is to try to alter pedestrian behavior. For this, society could also impose stricter laws and punishment on jaywalking. AVs could report such behavior to the local authorities as they are equipped with numerous sensors, including several cameras (Franke, Pfeiffer, Rabe, Knoeppel, Enzweiler, Stein and Herrtwich (2013); Ranft and Stiller (2016)). This would reduce the need for interaction as at crosswalks, at least in some countries, the vehicle (here the AV) has to yield. Colley et al. (2020a), however, already raised privacy concerns about AVs as monitoring devices. Nevertheless, such interactions will likely occur and have to be, therefore, accounted for.

8. Conclusion

With this study, it was shown that even with a clear introduction of AVs and its capabilities and responsibilities, pedestrians were still confused by the action of the human user within the vehicle. Operationalized via perceived safety, trust, cognitive load, willingness to cross, confusion statements and the final decision whether or not to cross, results showed that contradictory messages lead to high confusion. Also, the position of the human user played a moderating role in the assessment of the human user's role. Nevertheless, communication helped decide to cross. We showed that this confusion is lower in the intention-based communication and, therefore, add a supporting argument for introducing such communication as proposed in the latest ISO technical report on eHMIs (23049:2018 (2018)). Being mainly researched as a possibility to substitute current communication in traffic between human drivers and pedestrians, the evolving role as a mere passenger in an AV has to be considered. Thus, overall, this work helps to safely introduce AVs in general traffic.

ACKNOWLEDGMENTS

The authors thank all study participants. This work was conducted within the project 'Interaction between automated vehicles and vulnerable road users' (Intuitiver) funded by the Ministry of Science, Research and Arts of the State of Baden-Württemberg.

CRediT authorship contribution statement

Mark Colley: Conceptualization, Methodology, Software, Hardware, Investigation, Formal Analysis, Data Curation, Writing- Original Draft, Writing- Reviewing and Editing, Visualization, Project administration. **Bastian Wankmüller:** Software, Hardware, Investigation. **Tim Mend:** Software, Hardware, Investigation. **Thomas Väth:** Software, Hardware, Investigation. **Enrico Rukzio:** Resources, Supervision, Writing- Reviewing and Editing. **Jan Gugenheimer:** Conceptualization, Methodology, Writing- Reviewing and Editing.

References

23049:2018, I., 2018. Road Vehicles: Ergonomic Aspects of External Visual Communication from Automated Vehicles to Other Road Users. Standard. International Organization for Standardization.

Ackermans, S., Dey, D., Ruijten, P., Cuijpers, R.H., Pfleging, B., 2020. The effects of explicit intention communication, conspicuous sensors, and pedestrian attitude in interactions with automated vehicles, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi.org/10.1145/3313831.3376197, doi:10.1145/3313831.3376197.

author, U., 2020. lme4 convergence warnings: troubleshooting. https://rstudio-pubs-static.s3.amazonaws.com/33653_57fc7b8e5d484c909b615d8633c01d51.html. [Online; accessed 03-SEPTEMBER-2020].

Chai, Z., Nie, T., Becker, J., 2021. The Battle to Embrace the Trend. Springer Singapore, Singapore. pp. 179–249. URL: https://doi.org/10.1007/978-981-15-6728-5_7, doi:10.1007/978-981-15-6728-5_7.

Chang, C.M., Toda, K., Igarashi, T., Miyata, M., Kobayashi, Y., 2018. A video-based study comparing communication modalities between an autonomous car and a pedestrian, in: Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 104–109. URL: https://doi.org/10.1145/3239092.3265950, doi:10.1145/3239092.3265950.

Charisi, V., Habibovic, A., Andersson, J., Li, J., Evers, V., 2017. Children's views on identification and intention communication of self-driving vehicles, in: Proceedings of the 2017 Conference on Interaction Design and Children, Association for Computing Machinery, New York, NY, USA. p. 399–404. URL: https://doi.org/10.1145/3078072.3084300, doi:10.1145/3078072.3084300.

Colley, M., Mytilineos, S.C., Walch, M., Gugenheimer, J., Rukzio, E., 2020a. Evaluating highly automated trucks as signaling lights, in: Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM. Association for Computing Machinery, New York, NY, USA. p. 111–121. doi:10.1145/3409120.3410647.

Colley, M., Rukzio, R., 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, ACM. Association for Computing Machinery, New York, NY, USA. p. 212–222. URL: https://doi.org/10.1145/3409120.3410646, doi:10.1145/3409120.3410646.

Colley, M., Walch, M., Gugenheimer, J., Askari, A., Rukzio, E., 2020b. Towards inclusive external communication of autonomous vehicles for pedestrians with vision impairments, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi.org/10.1145/3313831.3376472, doi:10.1145/3313831.3376472.

Colley, M., Walch, M., Gugenheimer, J., Rukzio, E., 2019a. 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, Association for Computing Machinery, New York, NY, USA. p. 307–314. URL: https://doi.org/10.1145/3349263.3351521, doi:10.1145/3349263.3351521.

Colley, M., Walch, M., Rukzio, E., 2019b. For a better (simulated) world: Considerations for vr in external communication research, in: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, Association for Computing Machinery, New York, NY, USA. p. 442–449. URL: https://doi.org/10.1145/3349263.3351523, doi:10.1145/3349263.3351523.

Colley, M., Walch, M., Rukzio, E., 2020c. Unveiling the lack of scalability in research on external communication of autonomous vehicles, in: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–9. URL: https://doi.org/10.1145/3334480.3382865, doi:10.1145/3334480.3382865.

Cummings, M., Ryan, J., 2014. Point of view: who is in charge? the promises and pitfalls of driverless cars.

Cunningham, M., Regan, M.A., 2015. Autonomous vehicles: human factors issues and future research.

Deb, S., Carruth, D.W., Fuad, M., Stanley, L.M., Frey, D., 2020. Comparison of child and adult pedestrian perspectives of external features on autonomous vehicles using virtual reality experiment, in: Stanton, N. (Ed.), Advances in Human Factors of Transportation, Springer International Publishing, Cham. pp. 145–156.

Degani, A., Kirlik, A., 1995. Modes in human-automation interaction: Initial observations about a modeling approach, in: 1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century, IEEE. IEEE, New York, NY, USA. pp. 3443–3450.

Degani, A., Shafto, M., Kirlik, A., 1996. Modes in automated cockpits: Problems, data analysis and a modelling framework, in: ISRAEL ANNUAL CONFERENCE ON AEROSPACE SCIENCES, OMANUTH PRESS LTD. OMANUTH PRESS LTD, Tel Aviv, Haifa, Israel. pp. 258–266.

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., Terken, J., 2020. Taming the ehmi jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives 7, 100174.

Dey, D., Martens, M., Wang, C., Ros, F., Terken, J., 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, Association for Computing Machinery, New York, NY, USA. p. 82–86. URL: https://doi.org/10.1145/3239092.3265946, doi:10.1145/3239092.3265946.

Dey, D., Walker, F., Martens, M., Terken, J., 2019. Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external human-machine interfaces for automated vehicles, in: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 369–378. URL: https://doi.org/10.1145/3342197.3344523, doi:10.1145/3342197.3344523.

Edition, I., 2018. What Happened When Driver Put His Tesla on Auto Pilot? https://www.youtube.com/watch?v=kaAUHpeFj1c&t=53s. [Online; accessed 12-JUNE-2020].

Faas, S.M., Kao, A.C., Baumann, M., 2020a. 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, Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi.org/10.1145/3313831.3376484, doi:10.1145/3313831.3376484.

Faas, S.M., Mathis, L.A., Baumann, M., 2020b. External hmi for self-driving vehicles: Which information shall be displayed? Transportation Research Part F: Traffic Psychology and Behaviour 68, 171–186.

Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice 77, 167–181.

Franke, U., Pfeiffer, D., Rabe, C., Knoeppel, C., Enzweiler, M., Stein, F., Herrtwich, R.G., 2013. Making bertha see.

González, I.E., Wobbrock, J.O., Chau, D.H., Faulring, A., Myers, B.A., 2007. Eyes on the road, hands on the wheel: Thumb-based interaction techniques for input on steering wheels, in: Proceedings of Graphics Interface 2007, Association for Computing Machinery, New York, NY, USA. p. 95–102. URL: https://doi.org/10.1145/1268517.1268535, doi:10.1145/1268517.1268535.

Habibovic, A., Lundgren, V.M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S., et al., 2018. Communicating intent of automated vehicles to pedestrians. Frontiers in psychology 9, 1336.

Hart, S.G., Staveland, L.E., 1988. Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: Advances in psychology. Elsevier, Amsterdam, The Netherlands. volume 52, pp. 139–183.

Hecht, J., 2018. Lidar for self-driving cars. Optics and Photonics News 29, 26–33.

Hock, P., Benedikter, S., Gugenheimer, J., Rukzio, E., 2017. Carvr: Enabling in-car virtual reality entertainment, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 4034–4044. URL: https://doi.org/10.1145/3025453.3025665, doi:10.1145/3025453.3025665.

Holländer, K., Wintersberger, P., Butz, A., 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, Association for Computing Machinery, New York, NY, USA. p. 211–221. URL: https://doi.org/10.1145/3342197.3344528, doi:10.1145/3342197.3344528.

Hou, M., Mahadevan, K., Somanath, S., Sharlin, E., Oehlberg, L., 2020. Autonomous vehicle-cyclist interaction: Peril and promise, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–12. URL: https://doi.org/10.1145/3313831.3376884, doi:10.1145/3313831.3376884.

Inners, M., Kun, A.L., 2017. Beyond liability: Legal issues of human-machine interaction for automated vehicles, in: Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM. Association for Computing Machinery, New York, NY, USA. pp. 245–253.

Johnson, E.N., Pritchett, A.R., 1995. Experimental study of vertical flight path mode awareness. IFAC Proceedings Volumes 28, 153–158.

Joisten, P., Alexandi, E., Drews, R., Klassen, L., Petersohn, P., Pick, A., Schwindt, S., Abendroth, B., 2020. Displaying vehicle driving mode – effects on pedestrian behavior and perceived safety, in: Ahram, T., Karwowski, W., Pickl, S., Taiar, R. (Eds.), Human Systems Engineering and Design II, Springer International Publishing, Cham. pp. 250–256.

Joshi, A., Miller, S.P., Heimdahl, M.P., 2003. Mode confusion analysis of a flight guidance system using formal methods, in: Digital Avionics Systems Conference, 2003. DASC'03. The 22nd, IEEE. IEEE, New York, NY, USA. pp. 2–D.

Körber, M., 2019. Theoretical considerations and development of a questionnaire to measure trust in automation, in: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (Eds.), Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), Springer International Publishing, Cham. pp. 13–30.

Kothgassner, O.D., Felnhofer, A., Hauk, N., Kastenhofer, E., Gomm, J., Kryspin-Exner, I., 2013. Technology usage inventory (tui).

Kurpiers, C., Biebl, B., Mejia Hernandez, J., Raisch, F., 2020. Mode awareness and automated driving—what is it and how can it be measured? Information 11, 277.

Lankenau, A., 2001. Avoiding mode confusion in service-robots, in: Integration of Assistive Technology in the Information Age, Proc. of the 7th Int'l Conf. on Rehabilitation Robotics, IOS Press, Amsterdam, The Netherlands. pp. 162–167.

Leap, M., 2015. Magic Leap | Original Concept Video. https://www.youtube.com/watch?v=kPMHcanq0xM. [Online; accessed 12-JULY-2020].

Lee, S.H., Ahn, D.R., Yang, J.H., 2014. Mode confusion in driver interfaces for adaptive cruise control systems, in: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE. IEEE, New York, NY, USA. pp. 4105–4106.

Leveson, N., Pinnel, L.D., Sandys, S.D., Koga, S., Reese, J.D., 1997. Analyzing software specifications for mode confusion potential, in: Proceedings of a workshop on human error and system development, Glasgow Accident Analysis Group. Glasgow Accident Analysis Group, Glasgow, Scotland. pp. 132–146.

LLP, M.I., 2020. Autonomous/driverless car market - growth, trends, and forecast (2020 - 2025).

Löcken, A., Golling, C., Riener, A., 2019. How should automated vehicles interact with pedestrians? a comparative analysis of interaction concepts in virtual reality, in: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 262–274. URL: https://doi.org/10.1145/3342197.3344544, doi:10.1145/3342197.3344544.

LUMINQ, 2020. In-glass displays for improved automotive safety. https://www.lumineq.com/applications/automotive. [Online; accessed: 12-SEPTEMBER-2020].

Mahadevan, K., Somanath, S., Sharlin, E., 2018. Communicating awareness and intent in autonomous vehicle-pedestrian interaction, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–12. URL: https://doi.org/10.1145/3173574.3174003, doi:10.1145/3173574.3174003.

Matthews, M., Chowdhary, G., Kieson, E., 2017. Intent communication between autonomous vehicles and pedestrians.

Meme, J., 2020. Jins Meme. https://jins-meme.com/en/. [Online; accessed 12-SEPTEMBER-2020].

Mercedes-Benz, 2020. MBUX: Mercedes Benz User Experience. https://www.volkswagen.co.uk/technology/comfort/gesture-control. [Online; accessed 12-JULY-2020].

Millard-Ball, A., 2018. Pedestrians, autonomous vehicles, and cities. Journal of planning education and research 38, 6–12.

Moore, D., Currano, R., Strack, G.E., Sirkin, D., 2019. The case for implicit external human-machine interfaces for autonomous vehicles, in: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 295–307. URL: https://doi.org/10.1145/3342197.3345320, doi:10.1145/3342197.3345320.

Noguchi, K., Gel, Y.R., Brunner, E., Konietschke, F., 2012. nparld: an r software package for the nonparametric analysis of longitudinal data in factorial experiments. Journal of Statistical Software 50.

Norman, D.A., 1983. Design rules based on analyses of human error. Communications of the ACM 26, 254–258.

Pakusch, C., Stevens, G., Boden, A., Bossauer, P., 2018. Unintended effects of autonomous driving: A study on mobility preferences in the future. Sustainability 10. URL: https://www.mdpi.com/2071-1050/10/7/2404, doi:10.3390/su10072404.

Pfleging, B., Rang, M., Broy, N., 2016. Investigating user needs for non-driving-related activities during automated driving, in: Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia, Association for Computing Machinery, New York, NY, USA. p. 91–99. URL: https://doi.org/10.1145/3012709.3012735, doi:10.1145/3012709.3012735.

Pickering, C.A., Burnham, K.J., Richardson, M.J., 2007. A research study of hand gesture recognition technologies and applications for human vehicle interaction, in: 2007 3rd Institution of Engineering and Technology Conference on Automotive Electronics, IET. IET, Warwick, UK. pp. 1–15.

Qian, X., Ju, W., Sirkin, D.M., 2020. Aladdin's magic carpet: Navigation by in-air static hand gesture in autonomous vehicles. International Journal of Human–Computer Interaction 0, 1–16. URL: https://doi.org/10.1080/10447318.2020.1801225, doi:10.1080/10447318.2020.1801225, arXiv:https://doi.org/10.1080/10447318.2020.1801225.

Ranft, B., Stiller, C., 2016. The role of machine vision for intelligent vehicles. IEEE Transactions on Intelligent Vehicles 1, 8–19.

Rasouli, A., Kotseruba, I., Tsotsos, J.K., 2017. Understanding pedestrian behavior in complex traffic scenes. IEEE Transactions on Intelligent Vehicles 3, 61–70.

Rasouli, A., Tsotsos, J.K., 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems 21, 900–918.

Reifinger, S., Wallhoff, F., Ablassmeier, M., Poitschke, T., Rigoll, G., 2007. Static and dynamic hand-gesture recognition for augmented reality applications, in: Jacko, J.A. (Ed.), Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 728–737.

Rettenmaier, M., Pietsch, M., Schmidtler, J., Bengler, K., 2019. Passing through the bottleneck-the potential of external human-machine interfaces, in: 2019 IEEE Intelligent Vehicles Symposium (IV), IEEE. IEEE, New York, NY, USA. pp. 1687–1692.

Rettenmaier, M., Schulze, J., Bengler, K., 2020. How much space is required? effect of distance, content, and color on external human–machine interface size. Information 11, 346.

Riener, A., Ferscha, A., Bachmair, F., Hagmüller, P., Lemme, A., Muttenthaler, D., Pühringer, D., Rogner, H., Tappe, A., Weger, F., 2013. Standardization of the in-car gesture interaction space, in: Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 14–21. URL: https://doi.org/10.1145/2516540.2516544, doi:10.1145/2516540.2516544.

Rogers, K., Funke, J., Frommel, J., Stamm, S., Weber, M., 2019. Exploring interaction fidelity in virtual reality: Object manipulation and whole-body movements, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–14. URL: https://doi.org/10.1145/3290605.3300644, doi:10.1145/3290605.3300644.

Rosenthal, R., Cooper, H., Hedges, L., 1994. Parametric measures of effect size. The handbook of research synthesis 621, 231–244.

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., Ju, W., 2016. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles, in: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE. IEEE, New York, NY, USA. pp. 795–802.

Spencer Jr, C.F., 2000. Cockpit automation and mode confusion: The use of auditory inputs for error mitigation. Technical Report. AIR COMMAND AND STAFF COLL MAXWELL AFB AL.

Taxonomy, S., 2014. Definitions for terms related to on-road motor vehicle automated driving systems. Technical Report. Technical report, SAE International.

TeslaFi, 2020. TeslaFi Software Tracker. https://www.teslafi.com/firmware.php. [Online; accessed: 12-SEPTEMBER-2020].

Volkswagen, 2020. Gesture control. https://www.volkswagen.co.uk/technology/comfort/gesture-control. [Online; accessed 12-JULY-2020].

Walker, F., Dey, D., Martens, M., Pfleging, B., Eggen, B., Terken, J., 2019. Feeling-of-safety slider: Measuring pedestrian willingness to cross roads in field interactions with vehicles, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–6. URL: https://doi.org/10.1145/3290607.3312880, doi:10.1145/3290607.3312880.

Yusof, N.M., Karjanto, J., Terken, J., Delbressine, F., Hassan, M.Z., Rauterberg, M., 2016. The exploration of autonomous vehicle driving styles: Preferred longitudinal, lateral, and vertical accelerations, in: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Association for Computing Machinery, New York, NY, USA. p. 245–252. URL: https://doi.org/10.1145/3003715.3005455, doi:10.1145/3003715.3005455.

Mark Colley is a third-year PhD student at Ulm University in Germany. He is interested in the interaction with automation, in particular with automated vehicles. A main focus of his work lays in leveraging automation to support accessibility. Mark holds a Master's Degree in Computer Science from Ulm University. Contact him at mark.colley@uni-ulm.de

Bastian Wankmüller holds a Master's Degree in Media Computer Science from Ulm University. Contact him at bastian.wankmueller@uni-ulm.de

Tim Mend holds a Master's Degree in Computer Science from Ulm University. He is interested in software development especially in high performance computing. Contact him at tim.mend@uni-ulm.de

Thomas Väth holds a Master's Degree in Media Computer Science from Ulm University. He is interested in human-machine interaction, especially in context with automated vehicles and spoken dialogue systems. Contact him at thomas.vaeth@uni-ulm.de

Enrico Rukzio is currently a Full Professor with the Institute of Media Informatics, Ulm University, Ulm, Germany. His research interests include designing intelligent interactive systems that enable people to be more efficient, satisfied, and expressive in their daily lives. He received the Ph.D. degree in computer science from the University of Munich, Munich, Germany. Contact him at enrico.rukzio@uni-ulm.de

Jan Gugenheimer is currently an Assistant Professor in computer science with Telécom Paris (Institut Polytechnique de Paris), Paris, France, in the DIVA group. He is working on several topics around mixed reality (augmented reality and virtual reality) with focus on human–computer interaction. He received his Ph.D. degree in computer science from Ulm University, Ulm, Germany. Contact him at jan.gugenheimer@telecom-paris.fr