I See You! Design Factors for Supporting Pedestrian-AV Interaction at Crosswalks

Avram Block
Motional
Boston, Massachusetts, USA
aviblock@msn.com

Aryaman Pandya
Motional
Boston, Massachusetts, USA
aryaman.pandya@motional.com

Seonghee Lee
Cornell
Ithaca, New York, USA
seonghee.lee@motional.com

Paul Schmitt
MassRobotics
Boston, Massachusetts, USA
pauls@massrobotics.org

ABSTRACT

With the advent of autonomous vehicles (AVs) on public roads, the frequency of interactions between these AVs and pedestrians will increase. One example of such an interaction is at unsignalized crosswalks, where pedestrians and vehicles must negotiate for the right of way. Studies show that these interactions often use social communication channels. This paper addresses how AVs can fill this communication gap, focusing on the impact of pedestrian self-identifiability. Using VR, we designed two novel awareness-conveying behaviors, and a control condition with no awareness behavior. We then conducted a within-subjects VR study with 19 participants in which they traversed a crosswalk in front of a driverless vehicle in each experimental condition and rated their experience across seven probes. Results indicated that an awareness-conveying behavior significantly increased pedestrians' sense of safety and that increases in self-identifiability further improved pedestrians' experience without resulting in a heightened sense of surveillance from the vehicle.

CCS CONCEPTS

• **Human-centered computing** → Interaction design; Scenario-based design; Collaborative and social computing design and evaluation methods; • **Applied computing** → Transportation; • **Computer systems organization** → Robotics;

KEYWORDS

autonomous vehicles, pedestrian, HCI, eHMI design, social robotics

ACM Reference Format:
Avram Block, Aryaman Pandya, Seonghee Lee, and Paul Schmitt. 2023. I See You! Design Factors for Supporting Pedestrian-AV Interaction at Crosswalks. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23 Companion), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3568294.3580107

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9970-8/23/03…$15.00
https://doi.org/10.1145/3568294.3580107

1 INTRODUCTION

Traditional interactions between pedestrians and humans driving cars are rife with social exchanges, communicated through gesture, eye contact, and body language [7], [12]. These interactions signal to both parties that the other is aware of them and their intentions, and will plan accordingly. They are integral to the smooth functioning of societies in which pedestrians and vehicles operate in close proximity. One prime example of a site for this sort of interaction is at an unsignalized crosswalk, in which neither party receives explicit cues about who has the right of way. Instead, this question is answered by the collective judgement of those present.

As autonomous vehicles are launched onto public roads around the world, they are inevitably bound to find themselves in situations similar to those described above. However, the key communication channel that exists between pedestrian and human driver will be severed, and something must take its place. Significant research has been conducted on this topic, with many important and constructive findings. The dominant varieties of solutions include expressive movement dynamics of the AV [16] [3], audio cues [11], and most commonly, external graphical displays (eHMIs) [11], [5], [8]. While movement dynamics and auditory cues appear to be useful methods for supporting these interactions. Our work focuses on the use of external displays mounted on the AV.

Within the study of the use of external displays, there are multiple approaches that have been suggested. The primary options that have been identified within this space are to use either text-based [6], [15], graphical/icon-based displays [1], [9], or expressive animation of 1-dimensional LED light strips [4], [13]. Some have even studied the use of augmented reality for this purpose [17] Research suggests that graphical displays are the most effective of these three, in terms of legibility and visual salience. Thus, our work continues in this direction and explores factors to consider in the design of the graphical display to be presented on an eHMI for supporting pedestrian-AV interaction at crosswalks.

2 METHODS

The primary purpose of this study was to determine the overall viability of our design solutions to the problem of AV-pedestrian communication at road crossings, given the constraints suggested by the related work described above. Due to the logistically complex nature of our designs, we decided to avoid implementing and studying them using a physical prototype on a real AV. Instead, we

Figure 1: Depiction of each of the three experimental conditions. From left to right: C1 – Control Condition, C2 – Static Condition, C3 – Tracking Condition

conducted the experiment in Virtual Reality. This decision imposed limitations on the research as well, in terms of the complexity of the environment, and the immersiveness of the scenario. For these reasons, we relied on self-reported survey ratings as the primary output of the experiment. These survey questions were designed to characterize the participants' sense of safety and comfort throughout the scenario, with respect to the road crossing and the proximity to an AV.

2.1 Study Design

This study used a within-subjects design, in which each participant was presented with each of three experimental conditions exactly once, in counter-balanced random order. The three experimental conditions (shown in Figure 1) are as follows:

C1 - Control Condition: eHMI is not used, and remains off for the duration of the scenario

C2 - Static Condition: eHMI turns on when participant comes within "trigger" range (2m), and displays human-like figure which does not move at all throughout the scenario

C3 - Tracking Condition: eHMI turns on when participant comes within "trigger" range (2m), and displays human-like figure which moves according to pedestrian's relative position as they traverse crosswalk

2.2 Participants

For this study, we recruited a total of 19 participants, a mix of both internal (75%) and external (25%) participants with respect to employment at an AV company. Our sample population was 20% Female, 80% Male, and represented an age range from 22 - 50. A vast majority of participants had never experienced Virtual Reality before, while some had used it occasionally, and very few had used a VR headset semi-regularly.

2.3 Scenario Setup

2.3.1 VR Tools. In order to assess the viability of our proposed design, we decided to carry out this research in Virtual Reality, which allowed for more accurate representation of the intended designs than would have been possible with a prototype eHMI. Additionally, the technological development required for C3, in which the display follows the participant's movement across the crosswalk, is nearly negligible in VR, where the ground truth of the participant's location relative to the AV is always known. For the scenarios in this experiment, we worked with a 3D graphics consulting group to use Unreal Engine to design the environment, and to optimize for the Meta Quest 2 as our VR headset equipment.

2.3.2 Environment. Using these tools, we designed the virtual, experimental environment. Because this research focuses on minimizing pedestrian uncertainty in ambiguous roadway interactions with AVs, we chose an unsignalized four-way stop. The lack of traffic light in such an intersection requires right-of-way negotiation between the agents present. In our experimental environment, participants began by standing a few feet back from the curb at one corner of the intersection, facing across one of its entrances. At this closest entrance to the intersection, a generic next-generation AV sat stationary. This vehicle, pictured above contained multiple passengers, arranged such that the lack of a driver or driver's seat was visible upon inspection. It made varying use of its eHMI based on experimental condition, but did not physically move at any point in the study. This environmental setup was intended to replicate the common pedestrian experience of arriving at a four-way stop after the vehicle whose path intersects that of the pedestrian, thereby creating the need for a negotiation of which party will proceed in front of the other.

2.4 Procedure

2.4.1 Preliminary Stage. The procedure for this study began with the researcher explaining to the participant that the experiment revolves around pedestrian experience while crossing in front of a vehicle at an unsignalized roadway crossing. No explicit attempt was made to draw participants' attention to the eHMI on the vehicle, or to the fact that the vehicle was driverless. Participants were then asked preliminary rating questions, such as self-reported familiarity with autonomous vehicles, risk-propensity when crossing the street, and importance of their digital privacy.

2.4.2 Practice Stage. The VR procedure for this study made use of the Oculus Quest 2's motion tracking capabilities, and asked participants to physically walk 20 feet in a straight line while wearing the headset, in order to traverse the virtual crosswalk. Even for those from our sample who were more accustomed to Virtual Reality, this was a novel experience. Thus, the VR-based portion of the procedure began by presenting participants with a "practice" version of the environment, in which no vehicles were present. Participants

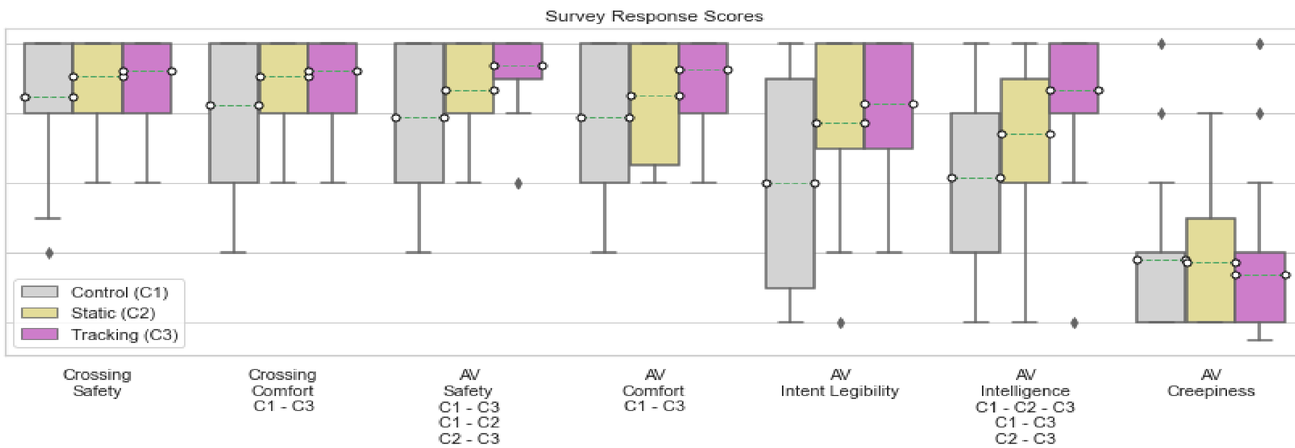

Figure 2: Survey Results Across Experimental Conditions and Probes. Dotted lines indicate mean values for each probe in each condition, and CX - CX labels beneath each probe indicate statistically significant differences in results.

were asked to walk back and forth across the virtual crosswalk until they felt comfortable with this interaction style between the virtual and physical worlds. The decision to use real-world walking for motion control in VR was intended to create a heightened sense of immersion, a higher level of realism with respect to pedestrian motion dynamics, and a lower sense of motion sickness than can often be experienced when moving through use of a joystick [14]. This practice session allowed participants to focus primarily on the virtual environment during the actual study, rather than being distracted by concerns about the safety of their movements in the real world.

2.4.3 Experimental Conditions. Once they felt comfortable in the practice environment, participants were presented with each of the experimental conditions. Regardless of the condition, participants were asked to place the headset on, orient to their virtual surroundings at the intersection, and walk towards their target location across the street. This required them to cross in front of the AV that was stopped at the intersection. After completing one traversal, they were asked to turn around and walk back to their starting position. During this entire process, participants were also asked to conduct a think-aloud exercise [19], in which they described everything they were thinking and witnessing while in the environment. These think-aloud narrations were recorded and transcribed for qualitative insights.

2.5 Measures

After experiencing each experimental condition, participants were asked a series of rating and Likert-style questions, all using a 5-point scale:

Likert:
L1: I felt safe crossing the street
L2: I felt comfortable crossing the street
L3: I felt safe around the vehicle in this scenario
L4: I felt comfortable around the vehicle in this scenario
L5: I understood the vehicle's intentions
Rating:
R1: Please rate the vehicle's intelligence on a scale from 1-5
R2: Please rate the vehicle's creepiness on a scale from 1-5

After having experienced each condition and responded to each query item, participants were invited to engage in a more informal, qualitative discussion of all three conditions, through the use of open-ended questions about each condition:
Q1: What did you like about this behavior?
Q2: What did you dislike about this behavior?
Q3: What would you change about this behavior?

2.6 Data Analysis

Likert and rating questions were collected across participants for each experimental condition. In order to determine significant trends in our survey response data, we conducted Wilcoxon tests between each pair of conditions for each question. In order to determine whether there was a significant difference between all three conditions, we used the Friedman test on each question. In both tests, we used a p-value of 0.05 as a threshold to assert significance.

3 RESULTS AND DISCUSSION

3.1 eHMI Benefit

Our results (displayed in Figure 2) on probes L1 and L2 ("I felt safe crossing the street", and "I felt comfortable crossing the street"), support the extensive body of pre-existing work that suggests that any use of an eHMI to communicate AV awareness at uncertain or unsignalized significantly improves pedestrians' subjective experience of these negotiations [11], [10]. There are significant differences between C1 and both C2 and C3, although no significant differences between C2 and C3 on this probe.

3.2 Static vs. Tracking Display

3.2.1 Continuous Feedback. Beyond this result, we also find significant improvements from C2 to C3 with respect to pedestrians' overall impression of the AV itself. This is reflected across probe results addressing pedestrians' sense of safety and comfort around the vehicle (L3 and L4), as well as their impression of the vehicle's intelligence. In considering our newfound data in conjunction with previous work done in this area, we believe that this difference may be due in part to the ways that C3, in which the human figure follows the pedestrian's location, addresses some of the concerns that previous researchers and users have expressed regarding static HMI displays such as C2. One of these concerns involves continued or repeated feedback over the course of the interaction between AV and pedestrian. Most eHMI solutions for supporting crossing decisions contain a static visual, which simply turns on when a pedestrian or AV arrives at the intersection, and remains on and constant until the end of the negotiation. This allows for a single instance of reactive communication from AV to pedestrian, but does not create a feedback loop which is responsive to the unfolding events of the road crossing. Previous research has shown the value of this feedback [5], [2], and comments from participants such as P15, who said of C3: "I like it because of the constant feedback", support that this aspect of the tracking condition provided valued feedback to pedestrians.

3.2.2 Recipient Ambiguity. The other well-documented concern with the use of eHMIs for this purpose is ambiguity when confronted with multiple pedestrians [5], [9], [18]. Our experiment was constrained to a single pedestrian, due to limitations with the use of VR for multiple simultaneous participants. However, the design put forth in C3 provides a more straightforward basis from which to address the problem posed by multiple pedestrians, and simultaneously assuages fears associated with this problem. With a static, binary display, such as presented in C2, it is unclear how an AV would effectively communicate recognition of a group of pedestrians, with different motion patterns and intentions. The state of the display being either on or off is intended to communicate with all those present at once. In fact, despite being alone in the VR scenarios, many participants immediately expressed concerns about C2's risk of causing confusion between multiple pedestrians, due to an inability to determine if the display was "meant" for them, or whether they had misinterpreted a signal intended for someone else in the scene. P9, when asked about his impression of C2, stated "I don't know what it's looking at. If it was looking at someone else, I wouldn't know and might falsely trust it."

On the other hand, qualitative results showed that C3 was a vast improvement over the static display of C2 in this vein. While performing the thinkaloud task during C3, many participants confidently recognized that the figure on the eHMI was moving along with their movements in the crosswalk, and noted that "that's me" that the car was perceiving. Further movement in the vicinity of the vehicle only served to confirm this suspicion, thereby removing any uncertainty about whether the display was, in fact, intended for them. In the course of this design research, we began to believe that this self-identifiability is a key component of an effective eHMI for pedestrian-AV interaction. The combination of the movement and humanlike form of the display in C3 led to high reports of self-identifiability, which coincided with elevated impressions of the AV's intelligence, and a heightened sense of comfort around the vehicle. While the problem of limited real estate in which to display multiple figures remains relevant, we propose that the use of larger, wraparound eHMI in conjunction with an appropriate "trigger" radius will minimize the impact of this issue.

When pursuing a higher degree of self-identifiability in the display, we also wanted to ensure that pedestrians' sense of privacy was not infringed upon. Thus, we asked participants to rate their impression of the AV's "creepiness" on a scale from one to five. Our hypothesis was that the increased sense of safety brought on by C3 might be accompanied by an increase in perceived creepiness. Instead, we found no significant differences in perceived creepiness across C1, C2, and C3. While we cannot conclude from this finding that self-identifiability is not correlated with loss of a sense of privacy, we do take it as a suggestion that there is more room to increase self-identifiability (discussed in Future Work) before these risks are realized and become detrimental to these kinds of interactions.

4 FUTURE WORK AND CONCLUSION

In the research described above, we conducted an experiment on the design of an eHMI display for autonomous vehicles. The use case inspected here was to signal to pedestrians that the AV is aware of their presence at a crosswalk, and that they are safe to proceed across. Through our research, we determined that one particularly important factor in designing for this task effectively is ensuring a high degree of "self-identifiability", that is, the ability for individual pedestrians to recognize themselves reflected on the eHMI display. In the most successful design put forth in this research, wee achieve a certain level of self-identifiability by using a human figure as the graphic, and by programming this figure to move along the eHMI in tandem with the intended pedestrian's movement across the crosswalk. We found that this design led to an increased sense of safety, comfort, and intelligence with respect to the autonomous vehicle.

Future extensions of this experimental design should increase the complexity of the scenario by introducing moving vehicles in order to raise the stakes of the negotiation, and by increasing the number of pedestrians present, in order to assess the scalability of proposed solutions. In future iterations of the designs presented herein, we take inspiration from some the feedback of some of our participants. There is clear potential to further increase the self-identifiability aspect of the designs by reflecting features such as a pedestrian's state (e.g., standing vs. walking), as well as more specific attributes of each pedestrian (e.g., gender, adult vs. child, ambulatory vs. wheelchair-using). These suggestions have the potential to address some of the concerns raised in our discussion, as they further decrease ambiguity for each individual pedestrian present in the vicinity of an AV.

Overall, we conclude that the use of eHMIs for assisting in pedestrian-AV social negotiation is a rich field of inquiry, with the potential to significant improve the public experience of integrating AVs into pedestrian-dense societies.

ACKNOWLEDGMENTS

This work is supported by leaders in the AV industry. We'd also like to thank our friend Karen Zhang, for getting this project off the ground, and Malte Jung for providing invaluable feedback.

REFERENCES

[1] Michael P. Clamann, Miles C. Aubert, and Mary L. Cummings. 2017. Evaluation of Vehicle-to-Pedestrian Communication Displays for Autonomous Vehicles.

[2] Mark Colley, Jan Henry Belz, and Enrico Rukzio. 2021. Investigating the Effects of Feedback Communication of Autonomous Vehicles. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Leeds, United Kingdom) *(AutomotiveUI '21)*. Association for Computing Machinery, New York, NY, USA, 263–273. https://doi.org/10.1145/3409118.3475133

[3] Hatice Şahin, Kevin Daudrich, Heiko Müller, and Susanne CJ Boll. 2021. Signaling Yielding Intent with EHMIs: The Timing Determines an Efficient Crossing. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Leeds, United Kingdom) *(AutomotiveUI '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 5–9. https://doi.org/10.1145/3473682.3480253

[4] Koen de Clercq, Andre Dietrich, Juan Pablo Núñez Velasco, Joost de Winter, and Riender Happee. 2019. External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. *Human Factors* 61, 8 (2019), 1353–1370. https://doi.org/10.1177/0018720819836343 arXiv:https://doi.org/10.1177/0018720819836343 PMID: 30912985.

[5] Debargha Dey, Kai Holländer, Melanie Berger, Berry Eggen, Marieke Martens, Bastian Pfleging, and Jacques Terken. 2020. Distance-Dependent EHMIs for the Interaction Between Automated Vehicles and Pedestrians. 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, 192–204. https://doi.org/10.1145/3409120.3410642

[6] Lex Fridman, Bruce Mehler, Lei Xia, Yangyang Yang, Laura Yvonne Facusse, and Bryan Reimer. 2017. To Walk or Not to Walk: Crowdsourced Assessment of External Vehicle-to-Pedestrian Displays. https://doi.org/10.48550/ARXIV.1707.02698

[7] Nicolas Guéguen, Sébastien Meineri, and Chloé Eyssartier. 2015. A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety Science* 75 (2015), 87–89. https://doi.org/10.1016/j.ssci.2015.01.018

[8] Kai Holländer, Ashley Colley, Christian Mai, Jonna Häkkilä, Florian Alt, and Bastian Pfleging. 2019. Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 27, 11 pages. https://doi.org/10.1145/3338286.3340138

[9] Philip Joisten, Ziyu Liu, Nina Theobald, Andreas Webler, and Bettina Abendroth. 2021. Communication of Automated Vehicles and Pedestrian Groups: An Intercultural Study on Pedestrians' Street Crossing Decisions. In *Proceedings of Mensch Und Computer 2021* (Ingolstadt, Germany) *(MuC '21)*. Association for Computing Machinery, New York, NY, USA, 49–53. https://doi.org/10.1145/3473856.3474004

[10] Stefanie M. Faas, Johannes Kraus, Alexander Schoenhals, and Martin Baumann. 2021. Calibrating Pedestrians' Trust in Automated Vehicles: Does an Intent Display in an External HMI Support Trust Calibration and Safe Crossing Behavior?. 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 157, 17 pages. https://doi.org/10.1145/3411764.3445738

[11] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Can Interfaces Facilitate Communication in Autonomous Vehicle-Pedestrian Interaction?. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 309–310. https://doi.org/10.1145/3173386.3176909

[12] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. 2017. Agreeing to Cross: How Drivers and Pedestrians Communicate. https://doi.org/10.48550/ARXIV.1702.03555

[13] Alexandria I. Rossi-Alvarez, Kevin Grove, Charlie Klauer, Melissa Miles, Andy Schaudt, and Zachary Doerzaph. 2022. Impact of Highly Automated Vehicle (L4/5 AV) External Communication on Other Road User Behavior. https://rosap.ntl.bts.gov/view/dot/64678 Tech Report.

[14] Dimitrios Saredakis, Ancret Szpak, Brandon Birckhead, Hannah A. D. Keage, Albert Rizzo, and Tobias Loetscher. 2020. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience* 14 (2020). https://doi.org/10.3389/fnhum.2020.00096

[15] Anna Schieben, Marc Wilbrink, Carmen Kettwich, Ruth Madigan, Tyron Louw, and Natasha Merat. 2019. Designing the Interaction of Automated Vehicles with Other Traffic Participants: Design Considerations Based on Human Needs and Expectations. *Cogn. Technol. Work* 21, 1 (feb 2019), 69–85. https://doi.org/10.1007/s10111-018-0521-z

[16] Paul Schmitt, Nicholas Britten, JiHyun Jeong, Amelia Coffey, Kevin Clark, Shweta Sunil Kothawade, Elena Corina Grigore, Adam Khaw, Christopher Konopka, Linh Pham, Kim Ryan, Christopher Schmitt, and Emilio Frazzoli. 2022. Can Cars Gesture? A Case for Expressive Behavior Within Autonomous Vehicle and Pedestrian Interactions. *IEEE Robotics and Automation Letters* 7, 2 (2022), 1416–1423. https://doi.org/10.1109/LRA.2021.3138161

[17] Wilbert Tabone, Yee Mun Lee, Natasha Merat, Riender Happee, and Joost de Winter. 2021. Towards Future Pedestrian-Vehicle Interactions: Introducing Theoretically-Supported AR Prototypes. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Leeds, United Kingdom) *(AutomotiveUI '21)*. Association for Computing Machinery, New York, NY, USA, 209–218. https://doi.org/10.1145/3409118.3475149

[18] Marc Wilbrink, Manja Nuttelmann, and Michael Oehl. 2021. Scaling up Automated Vehicles' EHMI Communication Designs to Interactions with Multiple Pedestrians – Putting EHMIs to the Test. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Leeds, United Kingdom) *(AutomotiveUI '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 119–122. https://doi.org/10.1145/3473682.3480277

[19] Xuesong Zhang and Adalberto L. Simeone. 2022. Using the Think Aloud Protocol in an Immersive Virtual Reality Evaluation of a Virtual Twin. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction* (Online, CA, USA) *(SUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 8 pages. https://doi.org/10.1145/3565970.3567706