This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

Deliverable 1.4

An extension of the human-factors methodological toolbox for human-AV interaction design research

Deliverable Status	**FINAL**
Type	**REPORT**
Dissemination level according to the Grant Agreement	**PUBLIC**

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

Deliverable 1.4 Extension of methodological toolbox for human-AV interaction design research

Project Information

Project Name:	SHAPE-IT – Supporting the interaction of Humans and Automated vehicles: Preparing for the Environment of Tomorrow
Grant Agreement:	860410
Project Duration:	1 October 2019 – 31 March 2024
Coordinator	Chalmers University of Technology

Document Information

Due Date	30Nov2023	**Due date in months**	M50
Date of submission	30Nov2023	**Month of submission**	M50

Status	**Author**	**Date**
Prepared by	Nikol Figalova, Naomi Mbelekani, Yue Yang, Liu Yuan-Cheng, Wilbert Tabone, Amir Hossein Kalantari, Ali Mohammadi, Xiaomi Yang	10Oct2023
Reviewed	Klaus Bengler, TUM	19 Nov. 2023
Approved	Jonas Bärgman, Chalmers	27 Nov 2023

Please cite as:

Figalova, N., Mbelekani, N.Y., Yang, Y., Cheng, Y.C., Tabone, W., Kalantari, A.H., Mohammadi, A., Yang, X. (2023). *An extension of the human-factors methodological toolbox for human-AV interaction design research.* (Deliverable D1.4 in the EC ITN project SHAPE-IT). SHAPE-IT Consortium. http://doi.org/10.17196/shape-it/2023/D1.4

Acknowledgements

This is a deliverable from the SHAPE-IT project – a project that is multi-disciplinary and with substantial interaction across the early-stage researchers (ESRs), supervisors across the six host universities, and the associated partners of SHAPE-IT (see list below). We consequently want to sincerely thank all individuals and organizations that have contributed to the SHAPE-IT project. We especially want to thank the associated partners, for making the secondments support the ESRs research, but also as a way for the ESRs to get insights into how other organizations work. Finally, this project would not have been possible without the Marie Skłodowska-Curie program. That is, this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410.

List of Associated Partners

Asymptotic AI	https://www.asymptotic.ai
BMW	https://www.bmw.com
Institute of Transport Economics (TØI)	https://www.toi.no
SWOV	https://www.swov.nl
Toyota Motor Europe NV/SA (TME)	https://www.toyota-europe.com
Viscando	https://viscando.com
Volkswagen	https://www.vw.com
Volvo Cars	https://www.volvo.com
Zenseact	https://www.zenseact.com

The information presented in this document represents opinions, findings, or perspectives that may not necessarily reflect the official stance or views of the associated partners, nor of the involved universities. The opinions, conclusions, or results provided herein are intended for informational purposes only and should not be interpreted as endorsements or official positions of any of the organizations involved.

Note 1: The term Early-Stage Researcher (ESR) is used extensively in this document. The ESRs are PhD candidates funded by the SHAPE-IT project.

Note 2: This document has been submitted to the EC for acceptance as a deliverable in the SHAPE-IT project. If changes are requested by the EC, the document will be updated.

Contents

List of Figures

List of Tables

List of Abbreviations

AV	Automated Vehicles
HMI	Human-Machine Interface
HUD	Head-Up Display
HMD	Head-Mounted Display
TTA	Time-to-Arrival
DTA	Difference in Time-to-Arrival
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
FOV	Field-of-View
ESR	Early State Researcher
VR	Virtual Reality
AR	Augmented Reality
EEG	Electroencephalogram
ECG	Electrocardiogram
ERP	Event-Related Potentials
AMICA	Adaptive Mixture Independent Component Analysis
TRASS	Transparency Assessment Method
EDA	Electrodermal Activity
eHMI	External-Human-Machine Interface
VRU	Vulnerable Road Users
HIKER	Highly Immersive Kinematic Experimental Research
DSS	Distributed Simulator Study

1 Introduction

Early state researchers (ESRs) of the SHAPE-IT project have committed to exploring innovative methods to ensure driving safety during interactions between human and Automated Vehicles (AVs). In this deliverable, insights of ESRs span a broad spectrum of methodologies, from experimental methods, including psychophysiological measures, Virtual Reality/Augmented Reality (VR/AR) applications, and transparency assessments, to human-AV interaction models, with vehicle-pedestrian model and vehicle-cyclist model, and lastly the long-term effects.

New types of interactions between humans and AVs need to be evaluated during the systems' development to ensure that requirements of safety, acceptance, and efficiency are met before they are introduced to the market. Since innovative concepts require great cost and effort for their realization, it is necessary to ascertain whether the expected effects will be achieved. Many of the systems' ergonomic requirements can be considered using experimental methods based on theoretical knowledge.

This proposal outlines different aspects for empirical investigations related to the interaction between human and AV. It is important to mention that different human roles need to be considered inside (passenger or driver) and outside/around (VRU) the AV. The research aspects range from cognitive processes (perception and decision), via motion behavior, to learning and behavioral adaptation. This requires that dedicated methods with clear, consistent definitions be refined or developed.

One example is the usage of virtual reality to investigate the complex interaction processes between AVs and VRUs in a safe and controllable setting as an alternative to field trials. Also, different AV communication strategies can be implemented in VR quicker and with reduced effort compared to hardware setups or experimental cars.

Further methods are physiological measurements, different types of driving simulation and long-term behavioral study approaches.

In their combination the different methods represent a toolbox of methodological approaches to analyze and evaluate different aspects of automated driving realizations.

This deliverable presents a collection of recommended experimental approaches that address complex questions using advanced measurement equipment and statistical approaches, and their successful application within the SHAPE-IT project.

2 Human-AV Interaction Design

Computer models and simulations alone cannot adequately test the acceptance and controllability of automated vehicles; more realistic experimental methods are necessary. Experimental tests can deliver the relevant data and promote a functional understanding to those who develop models of human behaviour that support interaction design. Furthermore, various requirements need to be fulfilled and ensured to guarantee safety. Thus, it’s necessary to develop a collection of methods specifically designed for verification and validation.

3 Methodology Developed

To construct a basis for future studies regarding automated vehicle interaction, ESRs in SHAPE-IT have developed various methods addressing different aspects of such interaction. In this section, these methods are roughly divided into three types: experimental methods (3.1.1 Neurophysiological Measures; 3.1.2 Transparency Assessments; 3.1.3 Augmented reality (AR) and virtual reality (VR)), modelling methods (3.2.1 Vehicle-pedestrian interaction at unsignalized locations; 3.2.2 Vehicle-cyclist interactions at unsignalized locations), and long-term methods (3.3 Researching Long-Term Effects of User-AV and VRU-AV Interactions).

3.1 Experimental Methods to Study Human-AV Interaction

3.1.1 Neurophysiological Measures

Driver-vehicle interactions can be assessed using different approaches. Self-report measures are easy to implement; however, they might be overly simplistic. Performance measures can provide good sensitivity for differentiating high-demand performance tasks; however, they are rather inefficient when studying low- to medium-demand tasks. Another possible approach is to use psychophysiological data, which is becoming more convenient due to the increasing availability of relatively cheap and precise wearable sensors. Psychophysiological measures such as electroencephalogram (EEG), electrocardiogram (ECG), and electrodermal activity (EDA) can be used to assess drivers' experience unobtrusively in real time. Disadvantages of psychophysiological measures include the requirements for special tools and equipment and the expertise to collect and analyse the data, and possibly a high signal-to-noise ratio (Kramer et al., 1995).

Electroencephalography (EEG) and event-related potentials (ERPs) are both non-invasive techniques used to record the electrical activity of the brain. The traditional approach to EEG uses an ongoing recording of the brain activity and the analysis of its spectral properties in the frequency domain. The ERP technique is different in that it averages the brain response to a specific event over many trials and produces results in the time domain. This has been shown to represent the postsynaptic potentials of pyramidal neurons and can provide insight into cognitive processes such as attention, perception, and decision-making.

The brain activity is measured on the scalp using electrodes and conductive gel. It offers very high temporal resolution (in milliseconds), but relatively low spatial resolution (i.e., areas of brain activities). The data can either be used for a frequency-domain EEG analysis or temporal-domain ERP analysis (for a review, see Lohani et al., 2019). In the field of driver-vehicle interaction, EEG and ERPs have been used to study how the brain processes information related to driving and how this information is used to guide the driver's behaviour.

For example, researchers have used EEG to study how drivers allocate their attentional resources when driving an automated vehicle (e.g., Van der Heiden, 2021, Figalová et al., in review) and as an objective measure of mental workload while drivers interact with an automated vehicle (e.g., Figalová et al., 2022, Lohani et al., 2021).

The versatility and non-invasive nature of EEG measurement makes it well suited for studying human cognition in driving contexts. These measures can be employed to study a wide range of questions, for example:

- How does driving a partially or fully automated vehicle affect our ability to pay attention to the road and manage our workload? EEG and ERP studies can help us understand the cognitive demands of different levels of automation.
- When driving an automated vehicle, what decision-making processes do we use, and how does our brain activity reflect those processes? EEG and ERP studies can help us understand the neural mechanisms of decision-making in automated vehicles.
- How do we perceive and respond to our environment while driving an automated vehicle, and what brain activity is associated with these processes? EEG and ERP studies can help us understand the cognitive processes involved in recognising traffic signs, other vehicles, and obstacles.
- How mentally drowsing is it to drive an automated vehicle, and how does our brain activity reflect the level of mental effort required? EEG can help us measure changes in brain activity associated with mental effort or fatigue during different stages of vehicle automation.
- How do distractions, such as mobile phone use or engaging with the vehicle's infotainment system, affect our ability to drive an automated vehicle, and what brain activity is associated with distracted driving? EEG and ERP studies can help us understand the neural mechanisms of driver distraction in automated vehicles.

ESR 1 and her colleagues studied the driver-vehicle interaction, evaluating both EEG and ERP. In our first experiment (Figalová et al., 2022), we assessed the effect of ambient visual cues conveying the current level of reliability of an automated vehicle on driver’s take-over performance. The experiment was performed in a fixed-base driving simulator (as in **Figure 1**).

Figure 1 Using EEG in a driving simulator at the Department of Human Factors, Ulm University, Germany

To record EEG, we used a standard actiCAP snap elastic cap with 32 Ag/AgCl active shielded electrodes placed according to the International 10-20 System with reference to FCz. The impedances were kept below 25 kOhm, and data were recorded with a sampling rate of 500 Hz using a LiveAmp amplifier and the BrainVision Recorder. MATLAB v2022a and EEGlab v2022.0 were used to pre-process the data, which consisted of high-pass filtering at 0.5 Hz, low-pass filtering at 30 Hz, bad channel removal, and channel interpolation. Independent component analysis was performed with the runica algorithm. We removed components that were flagged as originating primarily (more than 70%) from eye or muscular activity. The spectral characteristics were computed using the Darbelai package, with the frontal theta power calculated as a mean relative theta power on the Fz, F3, and F4 electrode locations. The parietal mean alpha power was calculated as a mean relative alpha power on the Pz, P3, and P4 electrode locations. The alpha and theta power were used as indicators of objective mental workload (Lohani et al., 2019), and the perceived mental workload was assessed using the Driver Activity Load Index (Pauzie et al., 2008).

The data suggest that there was no difference in the alpha and theta activity between drivers with and without the ambient light, indicating that the ambient light did not induce changes in objective mental workload. Moreover, no difference was found between the self-reported mental workload. Nevertheless, participants equipped with the ambient light demonstrated better performance, based on lower vehicle jerks. The results of this experiment show that ambient light helps drivers perform better after a take-over request.

Frequency domain analysis has, however, previously been criticised (Harpale et al., 2016). The results are sometimes inconsistent and heavily influenced by the pre-processing steps. Therefore, we stress the importance of documenting in detail how the data was handled, also in the pre-processing. Moreover, when researchers choose one of several possible approaches to pre-process the data, they should document why they chose this approach over the others.

A second experiment was performed in a prototype of an automated vehicle on a test track. In this experiment, ESR1 and colleagues focused on ERPs that were elicited by novel, task-irrelevant, environmental auditory cues.

Figure 2 Using EEG to record brain responses to novel environmental auditory cues in a prototype of an automated vehicle on a test track in Renningen, Germany

For the EEG recording in this experiment, we used the same ActiCAP set used in the first experiment (as in **Figure 2**). The data were digitised using the Lab Streaming Layer with a sampling rate of 1000 Hz. The EEG data were pre-processed in Matlab version R2022a with various plugins, including EEGlab v2022.1, ERPlab v9.00, IClabel v1.4, bemobil-pipeline 1.9, clean_rawdata v2.7, dipfit v4.3, xdfimport v1.18, and zapline-plus v1.2.1. We preprocessed the data using the BeMoBil pipeline and filtered the continuous EEG data with a band-pass filter between 0.1 and 100 Hz, removing power line artefacts using the ZapLine Plus plugin. We then interpolated bad channels and re-referenced each channel to the average of all channels. We high-pass filtered the data with a cut-off frequency of 1.25 Hz and then performed adaptive mixture independent component analysis (AMICA) with ten iterations. We enabled automatic rejection of bad data portions, keeping only the components that were most likely to originate from brain activity (as flagged by IClabel). After cleaning the raw data using the copied AMICA weights, we applied a second-order Butterworth filter and bandpass-filtered the data between 0.1 and 30 Hz. We then used continuous artefact rejection with a frequency range for thresholding between 20 and 30 Hz, setting the upper-frequency threshold to 10 dB with a 0.5-s epoch length. We required at least four contiguous epochs to label the region as artefactual, and we also expanded the region by 0.25 s on either side to ensure that the entire artefact was rejected.”. Finally, we extracted epochs from -200 to 800 ms after the stimulus onset, applied baseline correction, and rejected them if they contained step-like (artefactual) activity, using a threshold of 50 mV in a 200-ms moving window width with a 10-ms moving window step. The ESR1 and her colleagues replicated this approach in another experiment, conducted with a driving simulator instead of a test track. The results are currently being analysed.

Based on the experience gained in these experiments, we recommend considering the following when designing and conducting EEG/ERP experiments to evaluate the driver-vehicle interaction:

- **environmental noise**: real-world experiments will have more environmental noise than simulator experiments, so we recommend using larger sample size to have enough statistical power;
- **channel density**: high-density systems record the data with better spatial resolution, which allows better cleaning of the data using the independent component analysis;
- **sampling frequency**: 500Hz appeared to be sufficient; data with higher resolution was down-sampled before the analysis;
- **impedances**: given the mobile environment, the impedances should be kept as low as possible, generally speaking under 10 kOhm;
- **pilot trial:** we recommend scheduling several participants on different days before the beginning of the recruitment of participants for the actual experiment to ensure the hardware and software are working flawlessly;
- **data control**: after each pilot trial, we recommend opening the dataset and making sure that everything is recorded properly and the triggers are correct;
- **pilot data pre-processing**: we recommend pre-processing the trial data according to the planned pipeline, as certain mistakes are not visible before pre-processing;
- **audio-visual material**: researchers should collect sufficient audio-visual material of the trials for future presentations and publications;

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

- **laboratory consumables**: before every experiment, equipment should be tested to make sure everything works well and is ready to use (e.g., electrolyte gels tend to dry out between experiments);
- **documentation**: we recommend documenting every step in as much detail as possible in every stage of the experiment (planning, executing, data pre-processing, and analysis);
- **notes**: we recommend making notes about each participant, any non-standard situations, comments, etc.;
- **pre-registration**: pre-registering EEG/ERP studies is a useful way to avoid not being published due to the lack of a significant effect and mitigates bad scientific practice (such as data manipulation, p-hacking);
- **real-world testing**: whenever possible, test in a realistic environment instead of a driving simulator;
- **refreshment:** as EEG experiments require a lot of time, keep your participants hydrated and fed; always have bottles of water and snacks available and think about people with food allergies;
- **toilets:** suggest that your participants use the toilet after the EEG is mounted and before the experiment begins;
- **no pain induced:** make sure you stress to participants that the application of EEG gel can be uncomfortable, but not painful, and ask them to tell you whenever it hurts, in order to avoid participant complaints about bruises/scratches induced by the researcher;
- **two people setting up the EEG**: we recommend that two researchers set up the EEG, which expedites the process, provides a safeguard against any potential misconduct allegations, and creates a more comfortable and less intimidating environment for participants, particularly in scenarios involving mixed genders;
- **various times for set-up**: be prepared for the fact that the EEG set-up takes a different amount of time with each participant—we experienced times from seven to 80 minutes with the same 32-channel system;
- **buffer time**: following on the previous recommendation, many things can go wrong with an EEG set-up; therefore schedule some extra time between participants to

avoid delays and stress;

- **standardise the instructions**: we recommend presenting the conditions to participants using videos so everyone gets the exact same instructions, particularly when many researchers are collecting the data, to make sure that there is no bias;
- **online questionnaires:** whenever possible, choose online questionnaires over paper-and-pencil ones, because it saves time and minimises the risk of human error.

3.1.2 Transparency Assessments

In order to ensure driving safety during the automated ride, it is critical to assess whether human users understand the automated system. A human-machine interface (HMI), one of the few ways to transmit information from automated vehicles (AVs) to human users, is critically important and has been widely studied (Bengler et al., 2020). Researchers have developed various HMI design guidelines to meet the users' need for correct information and to minimise their workload (Young et al., 2017; Naujoks et al., 2019). However, these guidelines are either based on heuristics or subjective evaluations; to date, an objective and systematic assessment method does not exist. Creating such a method could increase the success of the HMI design process and the efficiency of the resulting designs.

To bridge the gap, ESR 7 developed the Transparency Assessment Method (TRASS), which can be used to objectively measure the understandability of the targeted AV HMI (Liu et al., 2022). In this study, "functional transparency" is defined as the combination of achieving a true (objective) understanding of the AV HMI and the workload required during that process. Functional transparency reflects the degree of ease in understanding the AV HMI. Moreover, it can be assessed objectively and numerically.

The TRASS was verified and used to find significant differences in the functional transparency of AV HMI designs in different automated driving systems (ADS). The study was conducted in the form of an online survey: each user was presented with ten scenarios and three AV HMI designs for each. Although no significant difference was found among the different AV HMI designs in functional transparency, further testing showed that certain design concepts might be the cause of the overall lower functional transparency of some AV HMI designs. It was concluded that the TRASS has the potential to improve the AV HMI development process, leading to designs that communicate quickly and correctly with users.

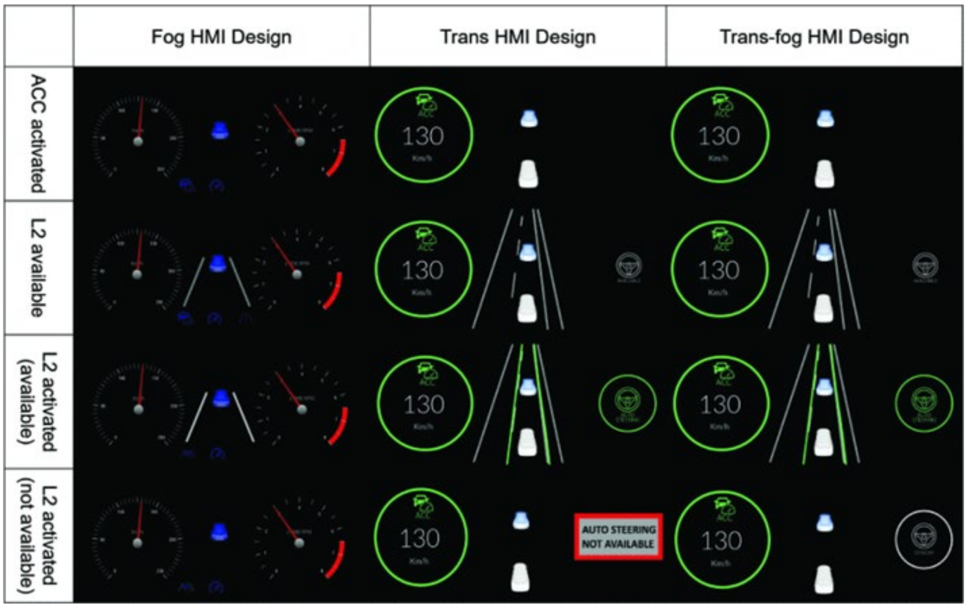

Figure 3 Human-machine interfaces designed and used for the studies.

To extend the scenario of the proposed method to a dynamic environment with higher fidelity, adaptation is necessary. In the online study, the workload of drivers trying to understand the AV HMI was quantified based on the time-to-understand of participants. However, a driving simulator or a test track would generate even more realistic real-time data, which cannot be obtained using the time-to-understand measure. A mental workload assessment method that can provide objective real-time data was required.

To determine which assessment methods are suitable for assessing the workload of AV HMI users while evaluating the designs, a study tested multiple psychophysiological measures (Liu et al., 2023). The study was conducted in a static driving simulator with a field of view of 120°, and the simulation software SILAB was used to create a 4-lane highway. The objective measures used included EEG, ECG, and electrodermal activity (EDA), which have been found to be sensitive in driving scenarios (Lohani et al., 2019). Results from the study found ECG and EDA are also sensitive and suitable for AV HMI design evaluation, as significant differences in these measures were found among AV HMI designs (as in **Figure 3**).

To summarise, ESR 7 proposed and validated a transparency assessment method, TRASS, which indicates to be effective in determining the degree to which users understood the AV HMI design. During the AV HMI design evaluation or design process, TRASS could serve as a powerful tool to assess, for instance, users' understanding of a newly designed HMI icon. In the future, TRASS could be used to identify the most transparent AV HMI designs based on the characteristics of human users, thus providing the safest possible environment during automated driving.

3.1.3 Augmented reality (AR) and virtual reality (VR)

Note that this part was included in this deliverable even if it relates to interactions with pedestrians outside of the AV, and the deliverable is part of the "inside AV" work package. We did this to provide a more complete presentation of toolboxes, and to demonstrate the multi-disciplinarity and multi-method aspects of SHAPE-IT.

Methods: Conceptualisation and design of novel AR interfaces for pedestrian-AV interactions

As part of the SHAPE-IT work package 2, nine novel AR interfaces for pedestrian-AV interactions were developed and presented in Tabone et al. (2021b) in a real crossing environment. The interfaces were designed using an experience-based ('genius') method which drew upon known theory and ideas extracted from expert interviews, which were published in Tabone et al. (2021a). These expert interviews brough forward the thoughts of 16 renowned human-factors experts on the future of automated vehicles, smart infrastructure, external-human-machine interfaces (eHMIs), and AR (specifically as a technology for communicating between AVs and vulnerable road users (VRUs)).

Brainstorming design sessions were organised to generate the final designs. The same scenario with a single vehicle approaching from the right on an unmarked road was envisaged for all the conceptualised interfaces. Each interface, to be displayed to the pedestrian, was designed to have two states, yielding and non-yielding, according to the vehicle's intention to stop (or not) for the pedestrian. These two states would be communicated to the pedestrian using green (yielding) and red (non-yielding). These colours were selected based on their high intuitiveness rating for signalling 'please cross/please do not cross', respectively (Bazilinskyy et al., 2020).

The online collaborative environment MIRO (Miro, 2023) was used as a supporting tool in the design process. An affinity diagram was used to collate ideas. A table with 15 design principles for human-computer interaction (Lee et al., 2017), LiDAR scans of the environment, and relevant theoretical literature were provided for reference. Following the affinity diagramming, a low-fidelity paper sketch was created and an AR heuristic evaluation (Endsley et al., 2017) was conducted for each AR interface concept, based on the design elements identified for that concept.

The interfaces were then implemented in AR for the iPad Pro 2020 model, using Unity MARS (Unity, 2023). Through this methodology, it was possible to create AR interfaces which could be adapted to the real-world test environment by setting semantic tags which represented various elements of the environment (e.g., the car position and the ground level). Each interface was then projected onto the real-world environment and filmed for demonstration

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

purposes. The final nine demoed AR interfaces are presented in **Figure 4**, and its caption provides a brief description of each.

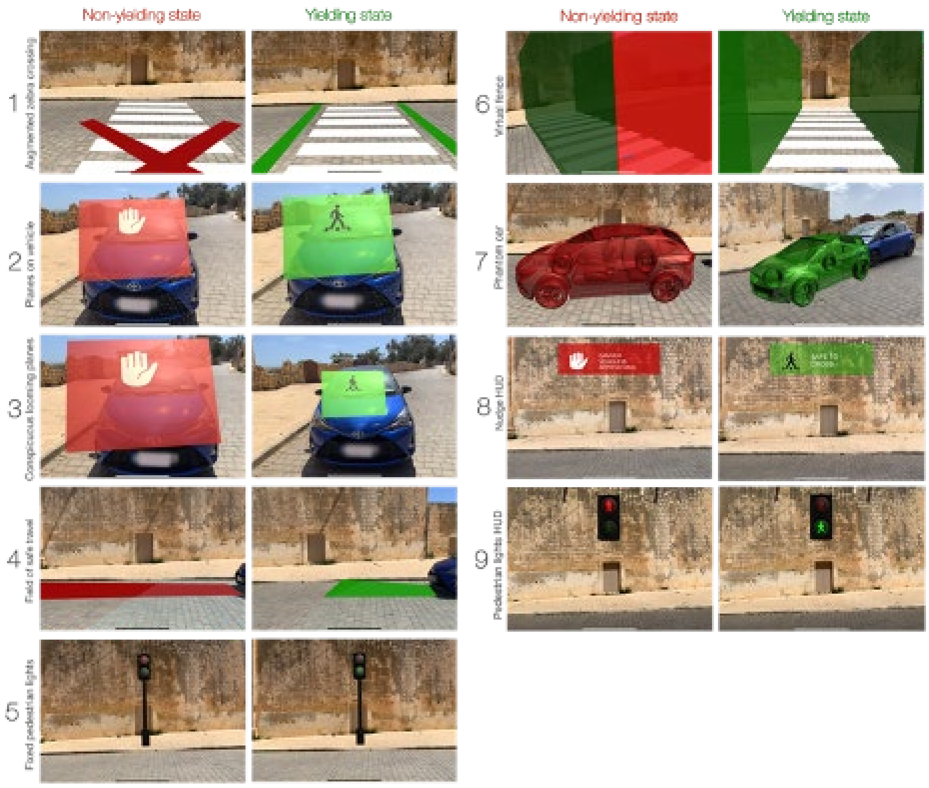

Figure 4 The nine AR concepts for pedestrian-vehicle interactions designed and developed by Tabone et al. (2021b) which are seen by the pedestrian. In total, nine AR interface concepts were developed, each with a yielding and non-yielding state: 1. Augmented zebra crossing, 2. Planes on vehicle, 3. Conspicuous looming planes (i.e., planes which grew or shrank in size), 4. Field of safe travel, 5. Fixed pedestrian traffic lights, 6. Virtual fence, 7. Phantom car (i.e., a transparent car which indicates the vehicle's predicted future position), 8. Nudge Head-up Display (HUD) (i.e., a floating text message and icon which informed the pedestrian whether or not it was safe to cross), 9. Pedestrian traffic lights HUD. Interfaces 1, 4, 5, 6, and 7 are projected on the road surface, while Interfaces 2 and 3 are projected on the car. Interfaces 8 and 9 are head-locked, i.e., they remain in the user's field of view. Image taken from Tabone et al., 2023.

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

Methods: Large-scale online questionnaire for the user evaluation of AR interfaces for pedestrian-AV interaction

Following the design process, a user evaluation was conducted using an online questionnaire which was successfully completed by 992 respondents in Germany, the Netherlands, Norway, Sweden, and the United Kingdom. Each interface was recreated in a VR environment (**Figure 4**) and presented displaying its yielding and non-yielding states in videos, which were presented in random order to the respondents.

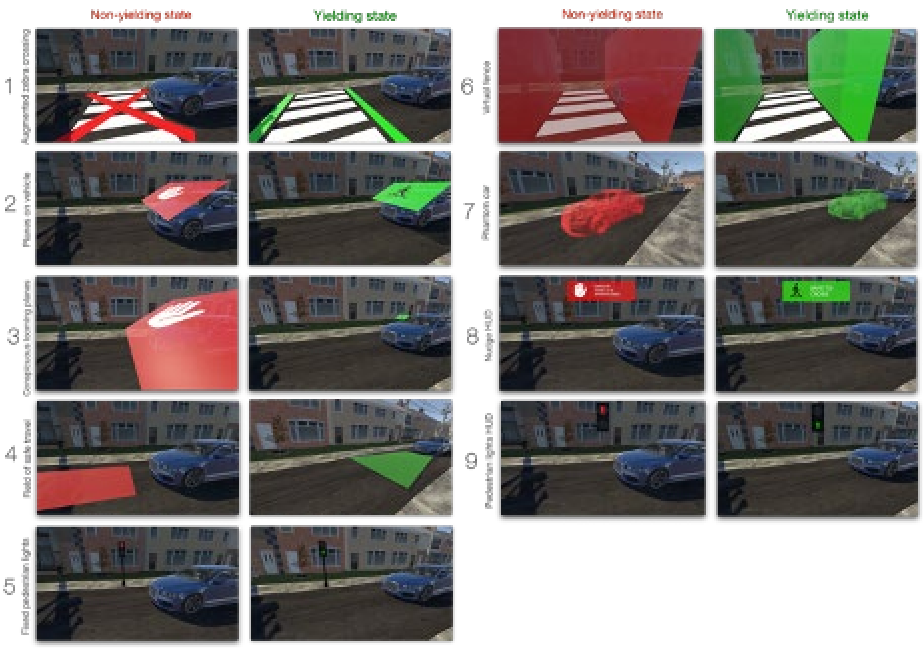

Figure 5 The nine AR concepts presented in the HIKER environment. Interfaces 1, 4, 5, 6, and 7 are projected on the road surface, while Interfaces 2 and 3 are projected on the car. Interfaces 8 and 9 are head-locked. Image taken from Tabone et al., 2023.

For each set of videos, the respondents were asked to rate the interfaces based on their intuitiveness, convincingness, aesthetics, and usefulness. Moreover, respondents were asked to provide free-text comments to further support their choices. While the method allowed for several conclusions to be made on the effectiveness of the designed AR interfaces and offered high repeatability, it did not offer high ecological validity. Further, it presented only low perceived risks to the participants. A possible solution to the problem of poor ecological validity was to use a VR simulation method, which embodies the participant within it.

Methods: CAVE pedestrian simulator for the user evaluation of AR interfaces for pedestrian-AV interactions

To raise the ecological validity of the method, a further experiment was conducted in the Highly Immersive Kinematic Experimental Research (HIKER), a CAVE pedestrian simulator located at the University of Leeds (University of Leeds, 2023). For this experiment, 30 participants with diverse backgrounds were recruited and asked to cross the virtual road when they felt safe to do so. The interfaces were presented in random order and states (as in **Figure 5**); a baseline condition with no interface was also included. Intuitiveness ratings and intention to cross were measured. The method included eye tracking, making this the first experiment in the HIKER to use such technology. Through the data provided by the eye tracker, it was also possible to analyse the attentional distribution of the participants. The aim of the method was to further analyse the pros and cons of each interface, and to present a way forward before testing in AR.

Methods: Extended Reality Head-Mounted Display for the user evaluation of AR interfaces for pedestrian-AV interaction

Additional methods were developed to analyse the interfaces in an VR (Peereboom, Tabone, et al., 2023), and AR Head-Mounted Display (HMD) (Aleva, Tabone et al., 2023). The VR HMD experiment allowed for the user evaluation of head-locked and world-locked AR interfaces, and conformal diminished reality (DR) designs, to assist pedestrians to cross in occluded scenarios. On the other hand, the AR HMD methods evaluated several interfaces developed in Tabone et al., 2021b in a real-world setting, where participants were able to experience the interfaces on an augmented vehicle that drove on a real road. In both experiments, the user trust and willingness to cross the road were measured.

Across all methods, various metrics such as intuitiveness, intention to cross, and trust were measured and compared among all designed AR interfaces. The combination of these methods provided a comprehensive evaluation methodology using AR, highlighting their effectiveness and areas for improvement.

3.2 Method to Model Human-AV Interaction

3.2.1 Vehicle-pedestrian interaction at unsignalized locations

3.2.1.1 Introduction

Vehicle-pedestrian interaction at uncontrolled locations has been a topic of interest in recent years. In addition to the safety aspect associated with vulnerable road users and the fact that pedestrians constitute a substantial proportion of all road users, recent developments in vehicle automation require a greater understanding of the details of their interaction with (automated) vehicles at locations with less clear regulations due to the unpredictable nature of human beings (FeldmanHall & Shenhav, 2019). To this end, both understanding (Amado et al., 2020) and modelling (Camara et al., 2020) vehicle-pedestrian interactions have become a necessity for the testing and development of highly automated vehicles (HAVs). Thus, ESR13 seeks to understand vehicle-pedestrian interaction and communication strategies by employing both conventional and behavioural game-theoretic models, using both lab and naturalistic data.

3.2.1.2 Methods

3.2.1.2.1 Experimental study

A distributed simulator study (DSS) was designed and conducted by connecting a high-fidelity motion-base driving simulator to a CAVE-based pedestrian lab, using an experimental paradigm in which two participants interact with each other in a safe and controlled environment. Thirty-two driver-pedestrian pairs participated together in 20 traffic scenarios. There were two crossing types (zebra and non-zebra) × five time-gaps (i.e., the temporal distance of the vehicle to the centre of the crossing; 3-7 s), with each repeated two times in two experimental blocks, resulting in 40 randomised trials per participant pair. In each traffic scenario, both participants were told to assume that they are in a hurry and can decide whether they want to pass the crossing first or wait for the other to do so (**Figure 6**). The pedestrians were prompted to step up to the kerb of the virtual road when they heard an auditory tone (which was triggered based on the temporal gap of the approaching vehicle) as in **Figure 7**. The speed limit of the two-way urban road was 30 mph and the drivers were encouraged to behave as they would during normal daily driving. Various objective metrics such as interaction outcome and pedestrians' and vehicles' kinematic features (e.g., trajectories), as well as subjective metrics such as the personality traits of both participants, were recorded and analysed (Kalantari, Yang, et al., 2023).

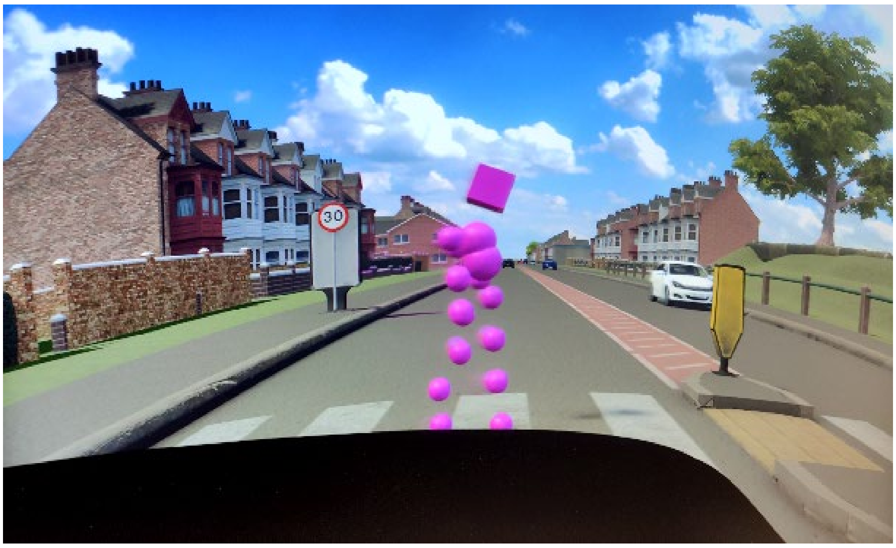

Figure 6 The driver's view of the pedestrian: the driver is stopping, and the pedestrian is represented by pink spheres.

Figure 7 the pedestrian's view of the vehicle in the pedestrian lab: the pedestrian is walking on the zebra crossing and the subject vehicle is to their right.

3.2.1.2.2 Naturalistic study

Real-traffic data was collected at two marked crossings in the city of Leeds. The choice of locations was based on several sessions of roadside observations and consultations with Leeds City Council regarding the safety and the prevalence of one-on-one vehicle-pedestrian interactions. The data were collected using two Viscando camera sensors over 14 days (with seven days at each location). The sensors detected road user types (i.e., light vehicle, heavy vehicle, cyclist, or pedestrian), tracking and recording their trajectory and speed over discretised time stamps as well as recording videos. An algorithm for detecting one-on-one interactions was developed and implemented, which extracted metrics like those in the experimental study from the processed data (Kalantari, Lin, et al., 2023)

3.2.1.2.3 Computational framework

Five computational models consisting of four game-theoretic and one logit model were proposed to predict the interaction outcomes in the studies. The framework consists of the following models: 1) A conventional game theory model [by Wu et al. (2019)], which considers the two-agent game of vehicle-pedestrian, was chosen and slightly modified. This model was established as the baseline for comparison against other models that employ more complex payoff formulations and solving algorithms. This model was named OCGT. 2) An alternative payoff formulation was introduced, rooted in Wu et al.'s original payoff. This formulation was provided to rectify certain assumptions made in the original payoff, which were suspected to not accurately capture the perceived utilities of different outcomes by road users. The model is named ACGT. 3) Both the original and alternative payoff formulations were handled using a model from the behavioural game theory category, leading to the creation of the OBGT and ABGT models, respectively. The dual accumulator model (Golman et al., 2020) was chosen and employed as an alternative game solution to the mixed-strategy Nash equilibrium. 4) A logit model, representing the popular modelling approach in this domain, was also tested and compared to the others. All models were simultaneously fitted to both crossing locations using maximum likelihood and Powell's method implemented in Scipy (Kalantari, Yang, Lee, et al., 2023).

3.2.1.3 Results

Figure 8 shows both agents' trajectories as a function of Time-to-Arrival (TTA) and crossing types for all trials. The interaction time, the time from the auditory tone to the time the vehicle passed the center of the crossing, is shown on the x-axis. From the figure, from left to right (as the time gap increases), the number of trials in which the pedestrian crossed first increases. It is also evident that pedestrians crossed first much more often when there was a zebra crossing (more green plots in the top five panels). Also, as expected, the drivers had to slow down or stop completely more often (more horizontal orange lines) when the pedestrian passed first. This difference is clear when comparing the four panels on the left (i.e., shorter time gaps) to the four on the right (i.e., longer time gaps) (Kalantari, Yang, et al., 2023).

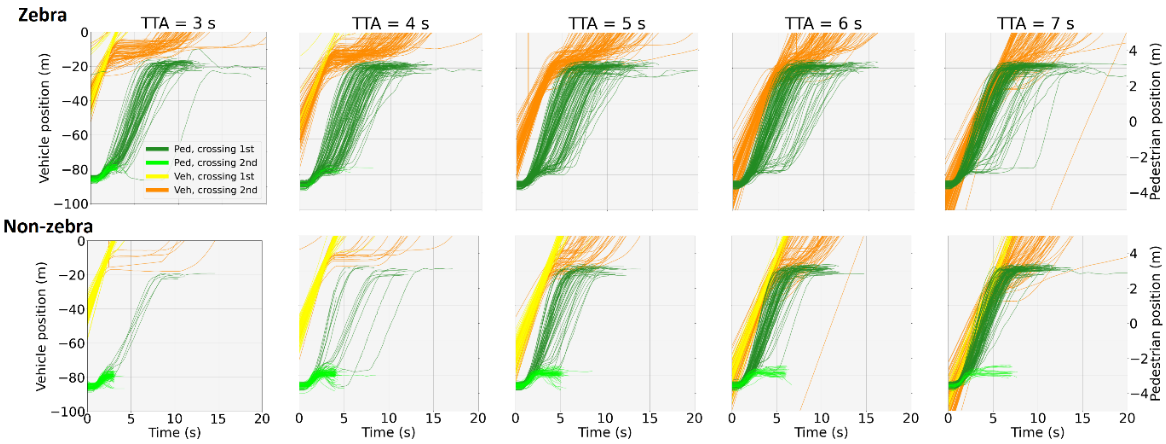

Figure 8 Vehicle and pedestrian trajectories over interaction time as a function of time gap and crossing type (Kalantari, Yang, et al., 2023).

Figure 9 shows the average of all 32 pedestrians' crossing probabilities over actual time gaps, for both crossing types and all models. The actual time gaps were calculated based on the designated time gaps minus one second to allow the pedestrians to have reached the kerb after hearing the auditory tone. Table 1 shows the model comparison for both crossing types including information loss criteria (Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC)) and error indices (Mean Absolute Error (MAE), Root Mean Square Error (RMSE)). Both from the figure and table, the proposed model of this project (ABGT) outperformed all other models regarding both crossing types. Also, the Wu et al. model combined with the dual accumulator model (OBGT) did a better job of capturing the pedestrian behaviour at both crossing types compared to the original model (OCGT). Finally, overall, the logit model performed well in predicting the interaction outcome, claiming the second-place for both crossings.

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

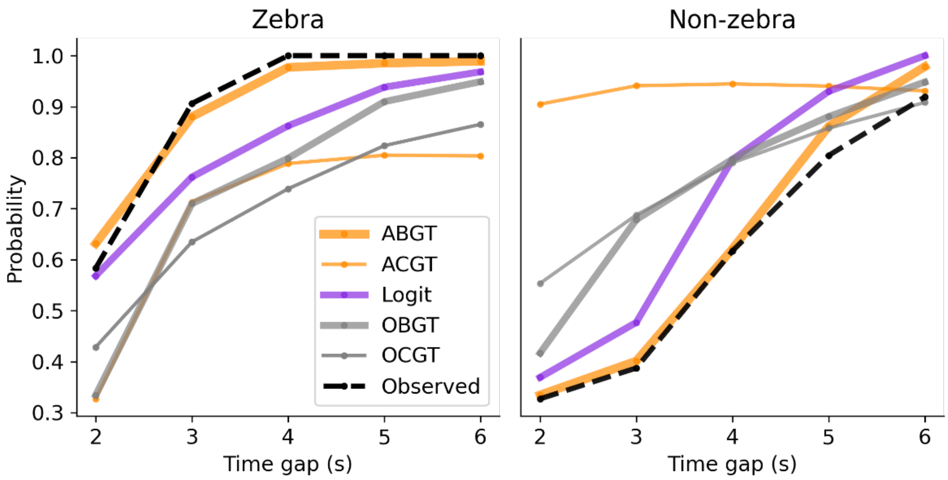

Figure 9 Average predicted probability of pedestrian crossing first over time gap for all models.

Table 1 Model comparison

Model	ABGTZ	ABGTNZ	ACGTZ	ACGTNZ	LogitZ	LogitNZ	OBGTZ	OBGTNZ	OCGTZ	OCGTNZ
MAE[1]	**0.0583**	**0.0878**	0.2121	0.4996	0.1226	0.1635	0.1725	0.2309	0.2315	0.2977
RMSE[2]	**0.0880**	**0.1393**	0.2372	0.5676	0.1497	0.1951	0.2097	0.2906	0.2627	0.3393
AIC[3]	**705.9494**		1984.9138		884.6076		1156.6956		1283.6192	
BIC[4]	**1540.8704**		2809.5272		1544.2983		1666.9251		1778.3872	
NLL[5]	**190.9747**		832.4569		314.3038		479.3478		545.8096	
NO param[6]	162		160		128		99		96	

1 Mean absolute error *2 Root mean squared error* *3 Akaike information criterion*

4 Bayesian information criterion *5 Negative log-likelihood* *6 Number of free parameters*

3.2.1.4 Conclusions and future work

This research project has three important findings. First, distributed simulations can generate scenarios where human road users can interact with each other and demonstrate behaviours like those observed in naturalistic studies. This experimental paradigm allows the study of many vehicle-pedestrian interactions in a safe and controlled manner, providing a powerful validation tool for computational models of road user interactions. Second, vehicle kinematics play a larger role in interaction outcomes than some psychological factors, such as the personality traits of the drivers and pedestrians at non-zebra crossings. Finally, drivers and pedestrians cannot be expected to act in line with the Nash equilibrium at unsignalized crossings, suggesting that more complex modelling paradigms such as behavioural game theory are needed to capture road user behaviours in these (and similar) scenarios. All these findings are beneficial for the future virtual testing of AVs before their deployment on roads.

Future work includes combining the computational models of this project with machine-learned ones to strike a balance between interpretability and generalisability and creating a structured approach for distributed simulations in VR, enabling multiagent interactions.

3.2.2 Vehicle-cyclist interactions at unsignalized locations

3.2.2.1 Introduction

At intersections, the intricate dynamics between vehicles and cyclists become particularly pronounced, spotlighting the need for comprehensive modelling and safety measures. The vulnerability of cyclists at intersections is underscored by escalating fatality rates among this group. When cyclists share space with motorised vehicles, conflicts arise due to differences in speed and visibility, as well as right-of-way interpretations. These interactions are further compounded at unsignalized intersections, where the absence of traffic control mechanisms intensifies the complexity of negotiations between cyclists and vehicles. The significance of modelling these interactions becomes evident in light of the need to decipher the nuanced behavioural patterns of both cyclists and drivers. By accurately capturing the variables influencing their decisions and actions, computational models enable the identification of potential collision scenarios and the development of effective countermeasures. Informed by these models, urban planners, traffic engineers, and vehicle manufacturers can collaboratively design safer infrastructure, optimise traffic management strategies, and engineer advanced driver assistance systems that prioritise cyclist safety. Thus, the modelling of vehicle-cyclist interactions at intersections emerges not only as a research imperative but a crucial link in forging a safer and harmonious coexistence between these distinct modes of transportation.

3.2.2.2 Method

3.2.2.2.1 Naturalistic study

Utilising stereovision and an AI-based sensor developed by VISCANDO, we gathered data from an unsignalized intersection in Gothenburg, Sweden. Over a span of 14 days, comprehensive information was captured regarding the movements of various road users, including pedestrians, cyclists, and light and heavy vehicles. The data, collected between 6:00 and 18:00 each day, comprised details such as vehicle positions, speeds, and headings, all sampled at a frequency of 20 Hz. This dataset was subsequently harnessed by SHAPE-IT to construct models elucidating the dynamics between cyclists and drivers. These models serve the dual purpose of offering a benchmark for autonomous vehicles (AVs) to emulate human drivers and furnishing predictive algorithms that empower AVs to decipher cyclists' intentions, thus contributing to enhanced interaction safety.

3.2.2.2.2 Experimental study

Data collection involved employing a bicycle simulator (***Figure 10***), on which participants traversed an unsignalized intersection multiple times, as per instructions. Through a virtual reality headset, participants were immersed in an environment representing an intersection featuring a car approaching from the cyclist’s right side. The task required participants to decide whether to cross before the car or yield and wait to cross after. Following the test, participants provided feedback on their simulator experience through a questionnaire. By amalgamating sensory data and responses to questionnaires, the researchers aimed to uncover the nuances of cyclists' interactions with autonomous vehicles (AVs) and the key

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

determinants influencing cyclists’ decisions. This study delved into the impact of varying factors such as the difference in time to arrival at the intersection (DTA) and field of view (FOV) distance on cyclists' decision-making processes.

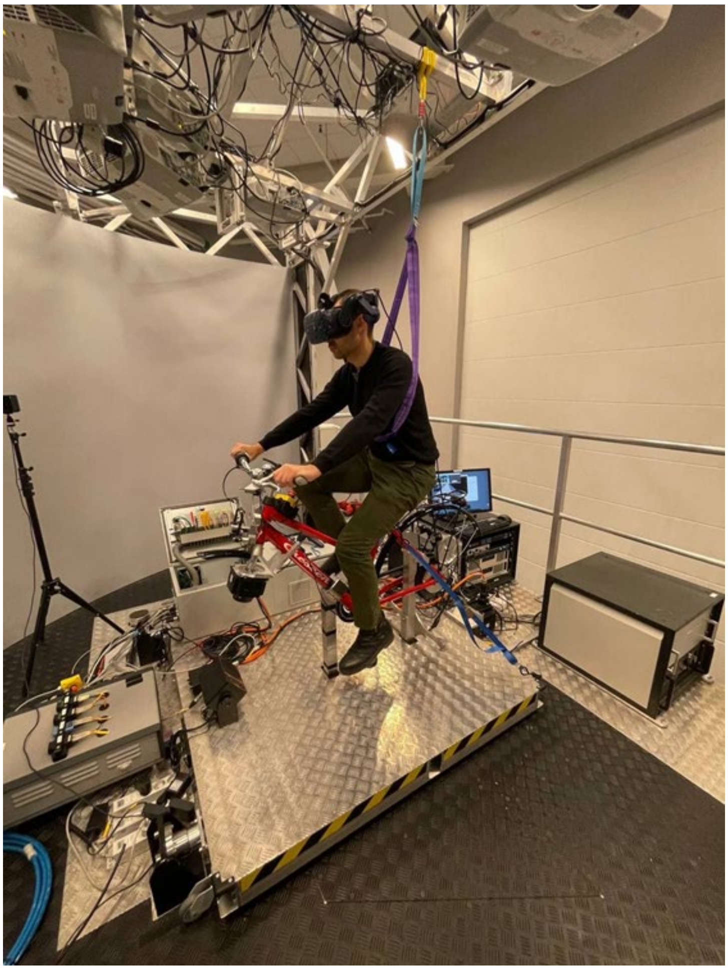

Figure 10 Riding simulator

3.2.2.3 Results

3.2.2.3.1 Naturalistic study

The estimation outcomes for the statistically significant variables concerning yielding probability are displayed in the table 2 below. It is noteworthy that due to the fewer yielding instances among cyclists compared to vehicles, we employed the Synthetic Minority Oversampling Technique (SMOTE; Chawla et al. 2002)) to redress the imbalance by oversampling. This action resulted in an augmentation in the number of observations to 136, evenly distributed between yielding cases for both cyclists and drivers.

Table 2 Summary of model estimate results

Variables	Coefficients	std err	P-value	lower bound (0.025)	Upper bound (0.975)
Intercept	-4.3523	1.474	0.003	-7.241	-1.464
Bike speed	-4.7794	2.041	0.019	-8.779	-0.780
Vehicle speed	9.4198	1.910	p < 0.001	5.676	13.163
DTA	5.5818	1.194	p < 0.001	3.242	7.922
Pedaling or not	1.1403	0.551	0.039	0.060	2.221
Looking or not	-1.4132	0.689	0.040	-2.765	-0.062

As depicted in ***Table 2***, the variables exerting significant influence on the decision to yield are: the cyclist's initial speed, the vehicle's initial speed, time to arrival (DTA), whether the cyclist was pedalling or not, and whether the cyclist was looking toward the vehicle or not. Each unit increment in the cyclist's speed translates to a 4.78 increase in the long odds of the cyclist crossing the intersection first. Conversely, an escalating vehicle speed heightens the likelihood that the vehicle will precede the cyclist through the intersection. The positive coefficient associated with DTA signifies that whichever road user is anticipated to reach the intersection earlier is more likely to go first. Furthermore, if the cyclist continues pedalling before the decision point, the likelihood of the cyclist crossing the intersection first increases by a factor of 3.12. Another significant parameter is the cyclist's gaze towards the approaching vehicle prior to the decision point; this orientation makes it 0.25 times more probable that the cyclist will traverse the intersection ahead of the vehicle.

3.2.2.3.2 Experimental study

This study delved into the decision-making process of cyclists on a riding simulator as they determined whether to cross an unsignalized intersection ahead of a vehicle. The investigation focused on the impact of two independent variables: difference in time to arrival (DTA) and field of view (FOV). Within the framework of a mixed-effects model, only DTA exhibited a significant effect on cyclists' crossing decisions. Although visibility didn't directly affect their choices, it did influence their approach speed to the intersection. Consequently, altering these independent variables across trials brought about changes in both bicycle kinematics and cyclist actions.

Throughout most trials, the cyclists' response pattern remained consistent, featuring actions like cessation of pedalling, braking, and attentive observation of the oncoming vehicle. The specific sequence of these actions varied based on the independent variables. These findings carry considerable importance, as a nuanced comprehension of cyclists' behaviour patterns serves as a cornerstone for crafting autonomous vehicles (AVs) that interact harmoniously and safely with cyclists. Notably, heightened visibility led to swifter cyclist reactions to approaching vehicles and smoother braking profiles, potentially mitigating the severity of their encounters with other vehicles. On the other hand, decreased DTA values correlated with a diminished likelihood that cyclists crossed first as well as with more abrupt bike braking profiles. These results provide invaluable insights towards establishing thresholds for safe DTAs that AVs can apply when crossing intersections and interacting with cyclists.

Another noteworthy discovery from the study, which is about the predictive significance of the cyclist’s glance toward the vehicle, underscores the significance of communication and eye contact between cyclists and drivers in shaping decision-making. As the realm of driverless vehicles expands, this aspect of human interaction will become less common, necessitating alternative forms of communication to ensure safety and foster comfort during cyclists’ interactions with automated vehicles.

The bike simulator employed in this study could be enhanced in the future, by incorporating cycling motion cues, an improved graphical interface, and increased lateral degrees of freedom. Overcoming the limitations of current bike simulators is pivotal, not only for recreating intricate cyclist interaction scenarios in virtual environments but also to avert the motion sickness experienced by some participants.

The outcomes illuminate how cyclists' kinematics and actions are influenced by DTA and FOV. This valuable information could lead to improved behavioural models capable of more accurately predicting cyclists' intentions during intersection crossings.

3.2.2.4 Conclusions and future work

In conclusion, this study has provided a comprehensive exploration of the decision-making process that cyclists undergo when faced with the choice of crossing an unsignalized intersection in the presence of a vehicle. Through meticulous utilisation of a riding simulator, the investigation strategically examined the influence of two pivotal independent variables, time to arrival (DTA) and field of view (FOV), on cyclists' crossing decisions. The results, analysed within the framework of a mixed-effects model, underscore the dominant role played by DTA in shaping cyclists' choices. While FOV didn't directly affect decision outcomes, its impact was evident in the cyclists' approach speeds. Crucially, the study's insights extend beyond individual decisions, delving into the consistent behavioural patterns of cyclists in response to approaching vehicles, offering a foundation for informed autonomous vehicle (AV) design. Enhanced visibility emerged as a key factor in promoting swifter cyclist reactions and smoother braking profiles, mitigating potential conflicts. Additionally, the study underscores the role of human interaction cues, such as communication and eye contact, which need to be reimagined as autonomous vehicles become prevalent. The study suggests avenues for refining simulation methods and emphasises the necessity of holistic approaches to model both cyclist and driver perspectives. Ultimately, this research significantly contributes to the understanding of cyclists' complex interactions with vehicles at intersections, laying the groundwork for improved safety measures and more accurate behavioural prediction models.

Moving forward, the focus will shift to explore the interactions between cyclists and various driver types at unsignalized intersections. Preliminary analysis based on the ND dataset indicates that expert drivers exhibited riskier interactions with cyclists compared to non-expert drivers. The forthcoming work aims to model cyclists' yielding decisions based on driver types, thus completing the holistic perspective of cyclist-vehicle interactions.

3.3 Researching Long-Term Effects of User-AV and VRU-AV Interactions

3.3.1 Introduction

As the deployment of highly automated vehicles (HAVs) and autonomous vehicles on real roads progresses, proactive research becomes imperative to anticipate their interactions with a wide range of user types and vulnerable road users (VRUs, including pedestrians, cyclists, and individuals with varying mobility challenges). The research must ensure that AVs can reliably detect, predict, and respond to different situations, as well as minimise the risk of accidents and injuries. However, there has not been much research that considers these factors over the long term. This neglect is since long-term research is one of the most critically challenging research approaches there is many scholars have argued that it is time-consuming, complex, and expensive, as well as lacking a precise definition (Mbelekani & Bengler, 2023). A major source of this imprecision is the difficulty in defining how long a period should be when considering potential changes in user behaviour (Mbelekani & Bengler, 2023). However, the development of automation technology that is safe in precarious situations depends on research that considers long-term effects of automation on both drivers and VRUs, on the road and in traffic. Consequently, there is a need to derive useful research strategies aimed at assessing long-term effects based on different interaction parameters and user experiences (e.g., trust, acceptance, transparency, comfort, etc.). The above-mentioned measuring methods (questionnaires, neurophysiological techniques, comfort assessments, and augmented and virtual reality) can all be used to assess long-term effects, in addition to models (or rather modelling) based on long-term data. These assessments will provide knowledge about the full spectrum of human factors and effects associated with human-

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

automation interactions (HAI) over time. The aim of this section is to provide useful insights that add to knowledge discovered in the field.

3.3.2 Methods

Long-term research helps predict and analyse how human-AV interactions (HAVI) might evolve over time, with special interest on user-AV as well as VRU-AV interaction patterns (see **Figure *11***). Understanding how humans (in this case, drivers, and pedestrians) adapt their behaviour towards AVs can guide the development of AV systems that anticipate and respond effectively to human behaviours, as well as responding safely in all scenarios. The guided development can inform the design of AVs' safety operation, communication strategies, and behaviours, enhancing user trust and acceptance. Additionally, insights from long-term research can contribute to infrastructure design.

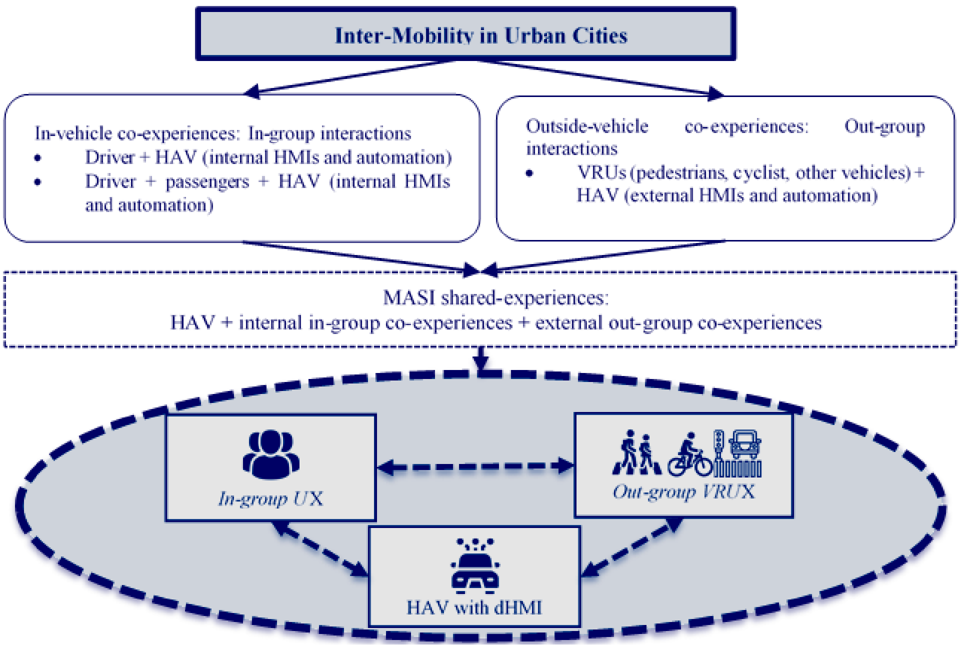

Figure 11 Inter-mobility and the intersection of in-group and out-group co-experiences (Mbelekani & Bengler, 2022).

Repeated measures are important for comprehensively understanding how human behaviour (specifically, how they interpret an AV's intent and adapt to its behaviour) can change over time. By increasing participants' exposures through repeated encounters, researchers can investigate how humans learn to interact with AVs, how their behaviours and perceptions evolve with increasing familiarity, and how these interactions can contribute (or not) to the overall safety and acceptance of AVs in future traffic eco system.

3.3.3 Assessing drivers' interaction with automation

Long-term assessments of user behaviour are important, as they may outline factors that are useful in properly understanding behaviour adaptation over time. Experimental driving simulator evaluations (see **Figure *12***) are regarded as beneficial as they are easier to control and manipulate. However, driving simulator studies have lower ecological validity than real-world studies, and for some types of experiments driving simulator studies are problematic (e.g., when studying the amplitude of braking, as the kinaesthetic feedback the driver gets in the driving simulator differs much from that in the real-world). As a result of the controlled setting, evaluations can be repeated precisely.

Figure 12 Automated driving simulation at the Chair of Ergonomics (TUM)

As a data-gathering tool, diaries are very useful; they allow the tracking of user experience over time. We employed a diary strategy in our study, with interviews both before and after. For the metrics, we chose consistent and informative evaluation tools and methodological approaches that demonstrably prove the users' understanding of interactive patterns and learning effects. Such evaluations are critical for advancing the underlying models, and for guiding designers and users to build wrongful expectations while reinforcing trust and acceptance. As Strömberg et al. (2018) stated, we need approaches that facilitate "flexible idea exploration, but in a contextualised and concrete manner through tangible objects and enactment to stage future use situations." Scholars are encouraged to envision and visualise new futures and new normalities, conceiving issues that do not yet exist (Mbelekani & Bengler, 2022). We need approaches that explore the imaginable future between users and AVs. In essence, we advocate for conventional and unconventional research strategies. We need approaches that explore the imaginable future between users and AVs.

3.3.4 Pedestrian simulation to investigate AV-VRU interaction

Given the limited presence of AVs in real-world scenarios, ESR 4 conducted a study using an advanced pedestrian simulation to explore the dynamics of AV-pedestrian interactions in long-term studies This research focused on the interaction between AVs and pedestrians. Especially, the introduction of AVs may introduce novel driving behaviour, communication strategies, such as external human-machine interfaces (eHMIs), that pedestrians may not be accustomed to or understand. This lack of familiarity has the potential to engender misunderstandings and confusion, leading to potentially hazardous situations. The Highly Immersive Kinematic Experimental Research (HIKER) simulator, a CAVE-based pedestrian simulator located at the University of Leeds, was used (**Figure *13***).

Figure 13 The HIKER pedestrian simulator.

Figure 14 Interaction between the pedestrian at the zebra crossing and the AV approaching from the right, with the eHMI on, in the HIKER lab. The yellow X indicates the pedestrian's starting position.

The CAVE-based environment offers precise control over the simulation, allowing researchers to manipulate various parameters to create a wide array of AV-pedestrian scenarios. This controlled experimentation provides insights into diverse interaction scenarios that may not be readily feasible in the real world, especially for the long-term studies. Meanwhile, simulation eliminates the safety concerns inherent in real-world AV interactions. Pedestrians can engage with AVs in a risk-free environment, which fosters exploration and candid behavioural responses. Pedestrian simulation emerges as a pivotal solution, offering a secure environment to address this challenge and evaluate the proposal of AV design strategies.

ESR 4 undertook this research. In the first experiment, the focus was on gathering data regarding pedestrians' crossing behaviour at a residential four-way crossing (**Figure *14***). This investigation encompassed the analysis of different approaching speeds of the AV, the presence/absence of an external Human-Machine Interface (eHMI), and the presence/absence of zebra crossings. These variables were meticulously examined to glean insights that could inform strategic AV design approaches. In a subsequent experiment, a distributed simulation setup was implemented: an intricate interplay was observed between pedestrians and drivers. This web of interactions was meticulously captured, yielding profound insights into pedestrians' behaviour as well as the design of AV's driving behaviour. For instance, pedestrians' head-turning behaviour is an important cue used by drivers to predict pedestrians' intention to initiate a crossing and their overall situation awareness (Rasouli et al., 2018). Using a VR environment, Yang et al. (2023) analysed pedestrians' head-turning behaviour during a range of road crossing scenarios involving AV interactions. The study showed that, as pedestrians' exposure to AVs without an external Human-Machine Interface (eHMI) display increased, their head-turning frequency decreased. However, when the AVs were equipped with an eHMI display, this pattern shifted markedly: the frequency of head-turning was significantly lower from the very first exposure but remained relatively constant as exposures increased. These findings underscore a promising avenue for future AV design that highlights the potential integration of an eHMI to enhance communication of the AV's intentions to pedestrians. The consistent head-turning behaviour in the presence of an eHMI suggests that such interfaces can effectively convey the AV's intent to pedestrians' right from the outset. However, the limited variation in head-turning behaviour with repeated exposures could indicate that pedestrians may not significantly adapt or learn over time when an eHMI is present, highlighting the need for further research on long-term interactions and potential habituation effects.

The pedestrian crossing rate measures the likelihood of the pedestrian's decision to initiate a crossing in front of the vehicle. The crossing rate was analysed in an experiment by Kalantari et al. (2022). The results show that as pedestrians' exposures to vehicles increased, their crossing rates exhibited a distinct increase. This finding was most notable for time gaps of 3 s or 4 s. However, when the time gap increased to 5 s, the crossing rate no longer changed with increasing exposures. This finding shows that the behaviour of pedestrians undergoes a transformation with increasing exposures and familiarity with the vehicle. Yet, the time gaps, encompassing vehicular kinematics, such as vehicle's deceleration and lateral deviation remain pivotal influencers. Further results are currently being analysed, with a particular focus on understanding how these kinematic cues influence pedestrians' decisions to cross.

Additionally, future research can investigate possible changes in other behaviours over time, such as gaze patterns, crossing initiation time, and crossing speed. Ultimately, these investigations may contribute to the future design of AVs, fostering more trustworthy driving behaviour and enhanced eHMIs.

3.3.5 Final remarks

The overall design philosophy requires automation to be capable of coordination, cooperation, and collaboration with users in different interactive tasks. Applying this philosophy means that effective research strategies between humans and automation need to be considered. Research into long-term effects may provide insights that help industry resolve challenging issues in different interaction scenarios, ultimately fostering safe and risk-free HAI architectures. Overall, by tracking humans' behaviours and decisions across multiple AV encounters, researchers can capture the nuances of learning and adaptation that occur as humans gain experience with AVs. This approach enables the exploration of trends, patterns, and potential inflection points in human responses to AV behaviour. As AV technology advances and interactions become more common, methodologies examining long-term effects will become crucial. This research method promises to contribute to the overall successful integration of AV technology into urban environments.

4 Conclusion

This deliverable provides insights into both methodologies developed in the SHAPE-IT project (incl. benefits and challenges), and uses of a variety of existing methods, across several of the SHAPE-IT ESRs. For example, some ESRs provide insights and recommendations for methodologies used to explore human-AV interaction. For EEG/ERP experiments, multiple design and methodological recommendations were listed in the part describing the neurophysiological measures. Furthermore, this report also introduces the Transparency Assessment Method (TRASS) for AV HMI design evaluations. The TRASS method has been validated and can be used to identify the most transparent AV HMI designs, based on user characteristics. Usage of AR/VR in human-AV interaction studies was also explored. Experiments under different test environments were conducted in SHAPE-IT, and the effects of different interface designs using AR/VR were identified, using a variety of methods.

This report also explored decision-making process models of cyclists and pedestrians when crossing an unsignalised intersection – exploring different methods for modelling interactions between cars, and pedestrians and cyclists. The results underscored the dominant role played by the difference in time to arrival (DTA) in shaping cyclists' decision whether to yield. The results of pedestrian-vehicle interactions show that distributed simulation can provide scenarios in which human road users show behaviours like those observed in naturalistic studies. The study also found that vehicle kinematics play a large role.

Finally, the study emphasises the importance of long-term research in resolving challenging issues in different interaction scenarios, to construct a safe and risk-free environment. The research method holds promise for shedding light on how humans learn to interact with AVs and adapt their behaviours, thus contributing to the overall successful integration of AV technology into urban environments.

5 References

Amado, H., Ferreira, S., Tavares, J. P., Ribeiro, P., & Freitas, E. (2020). Pedestrian–vehicle interaction at unsignalized crosswalks: A systematic review. *Sustainability*, *12*(7), 2805.

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2020). External Human-Machine Interfaces: Which of 729 colors is best for signaling 'Please (do not) cross'? *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 3721–3728, Toronto, Canada. https://doi.org/10.1109/SMC42975.2020.9282998

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI framework for automated driving. Information, 11(2), 61.

Camara, F., Bellotto, N., Cosar, S., Weber, F., Nathanael, D., Althoff, M., Wu, J., Ruenz, J., Dietrich, A., & Markkula, G. (2020). Pedestrian models for autonomous driving part ii: High-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, *22*(9), 5453–5472.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357

Endsley, T. C., Sprehn, K. A., Brill, R. M., Ryan, K. J., Vincent, E. C., & Martin, J. M. (2017, September). Augmented reality design heuristics: Designing for dynamic interactions. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 61, No. 1, pp. 2100-2104). Sage CA: Los Angeles, CA: Sage Publications. https://doi.org/10.1177/1541931213602007

FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behaviour*, *3*(5), 426–435.

Figalová, N., Chuang, L. L., Pichen, J., Baumann, M., & Pollatos, O. (2022). Ambient light conveying reliability improves drivers' takeover performance without increasing mental workload. *Multimodal Technologies and Interaction*, *6*(9), 73.

Figalova, N., Bieg, H. J., Reiser, J. E., Baumann, M., Chuang, L., & Pollatos, O. (in review). Attenuated response to environmental stimuli in automated driving: Multiple resources or multimodal integration?

Golman, R., Bhatia, S., & Kane, P. B. (2020). The dual accumulator model of strategic deliberation and decision making. *Psychological Review*, *127*(4), 477.

Harpale, V. K., & Bairagi, V. K. (2016, January). Time and frequency domain analysis of EEG signals for seizure detection: A review. In 2016 International Conference on Microelectronics, Computing and Communications (MicroCom) (pp. 1-6). IEEE.

Kalantari, A. H., Lin, Y.S., Mohammadi, A., Merat, N., & Markkula, G. (2023). Investigating driver-pedestrian interactions at marked crossings: A comparison of two methodologies. DOI: https://doi.org/10.31234/osf.io/gk9af

Kalantari, A. H., Yang, Y., Merat, N., Lee, Y.M., & Markkula, G. (2023). Driver-Pedestrian Interactions at Unsignalized Crossings Are Not in Line With the Nash Equilibrium. IEEE Access, 11, 110707-110723 https://doi.org/10.1109/ACCESS.2023.3322959

Kalantari, A. H., Yang, Y., Pedro, J. G. de, Lee, Y. M., Horrobin, A., Solernou, A., Holmes, C., Merat, N., & Markkula, G. (2023*). Who goes first? A distributed simulator study of vehicle–pedestrian interaction*. Accident Analysis & Prevention, 186, 107050. https://doi.org/10.1016/j.aap.2023.107050

Kramer, A. F., Trejo, L. J., and Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. Biological psychology 40, 83–100

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). Designing for people: An introduction to human factors engineering.

Liu, Y.-C., Figalová, N., & Bengler, K. (2022). Transparency Assessment on Level 2 Automated Vehicle HMIs. Information, 13(10), 489. MDPI AG. Retrieved from http://dx.doi.org/10.3390/info13100489

Liu, Y. C., Figalova, N., Baumann, M., & Bengler, K. (2023, June). Human-Machine Interface Evaluation Using EEG in Driving Simulator. In 2023 IEEE Intelligent Vehicles Symposium (IV) (pp. 1-6). IEEE.

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. Frontiers in human neuroscience, 13, 57.

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., et al. (2021). No difference in arousal or cognitive demands between manual and partially automated driving: a multi-method on-road study. Frontiers in neuroscience , 627

Mbelekani, N.Y. & Bengler, K. Systemizing Long-Term Research: Assessing Long-Term Automation Effects and Behaviour Modification. (2023, in press)

Mbelekani, N. Y., & Bengler, K. (2022, June). Traffic and Transport Ergonomics on Long Term Multi-Agent Social Interactions: A Road User's Tale. In International Conference on Human-Computer Interaction (pp. 499-518). Cham: Springer Nature Switzerland.

Miro. (2023). An Online Visual Collaborative Platform for Teamwork. https://miro.com/

Naujoks, F., Wiedemann, K., Schömig, N., Hergeth, S., & Keinath, A. (2019). Towards guidelines and verification methods for automated vehicle HMIs. Transportation research part F: traffic psychology and behaviour, 60, 121-136.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, 2(4), 315-322.

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2018). Understanding Pedestrian Behavior in Complex Traffic Scenes. IEEE Transactions on Intelligent Vehicles, 3(1), 61–70. https://doi.org/10.1109/TIV.2017.2788193

SAE International. Taxonomy and Definitions for Terms Related to Driving Automation Systems for on-Road Motor Vehicles (J3016_202104); SAE International: Warrendale, PA, USA, 2021.

Strömberg, H., Pettersson, I., Andersson, J., Rydström, A., Dey, D., Klingegård, M., & Forl zi, J. (2018). Designing for social experiences with and within autonomous vehicles – exploring methodological directions. Design Science, 4, E13. https://doi.org/10.1017/dsj.2018.9

Tabone, W., De Winter, J. C. F., Ackermann, C., Bärgman, J., Baumann, M., Deb, S., Emmenegger, C., Habibovic, A., Hagenzieker, M., Hancock, P. A., Happee, R., Krems, J., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B., & Stanton, N. A. (2021a). Vulnerable road users and the coming wave of automated vehicles: Expert perspectives. *Transportation Research Interdisciplinary Perspectives, 9*, 100293. https://doi.org/10.1016/j.trip.2020.100293

Tabone, W., Lee, Y. M., Merat, N., Happee, R., & De Winter J. C. F. (2021b). Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes. *Proceedings of the 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 209–218). https://doi.org/10.1145/3409118.3475149 ,

Tabone, W., Happee, R., García, J., Lee, Y. M., Lupetti, L., Merat, N., & De Winter J. C. F. (2023). Augmented reality interfaces for pedestrian-vehicle interactions: An online study. *Transportation Research Part F: Traffic Psychology and Behaviour, 94,* 170-189. https://doi.org/10.1016/j.trf.2023.02.005

Unity. (2023). Unity MARS Overview. https://docs.unity3d.com/Packages/com.unity.mars@1.5/manual/index.html

University of Leeds. (2023). Highly Immersive Kinematic Experimental Research (HIKER) pedestrian lab. https://uolds.leeds.ac.uk/facility/hiker-lab

Van der Heiden, R. M., Kenemans, J. L., Donker, S. F., and Janssen, C. P. (2021). The effect of cognitive load on auditory susceptibility during automated driving. Human factors, 0018720821998850

Wu, W., Chen, R., Jia, H., Li, Y., & Liang, Z. (2019). Game theory modeling for vehicle–pedestrian interactions and simulation based on cellular automata. *International Journal of Modern Physics C*, *30*(04), 1950025.

Yang, Y., Lee, Y., Madigan, R., Solernou, A., & Merat, N. (2023). Interpreting pedestrians' head movements when encountering automated vehicles at a virtual crossroad. Retrieved from https://osf.io/preprints/psyarxiv/n98v2

Young, K. L., Koppel, S., & Charlton, J. L. (2017). Toward best practice in Human Machine Interface design for older drivers: A review of current design guidelines. Accident Analysis & Prevention, 106, 460-467.