Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment

April 2023

Wilbert Tabone [1], Riender Happee [1], Yue Yang [2], Ehsan Sadraei [2], Jorge García [2], Yee Mun Lee [2], Natasha Merat [2], Joost de Winter [1]

[1] Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, The Netherlands
[2] Institute for Transport Studies, Faculty of Environment, University of Leeds, The United Kingdom

Abstract

Augmented reality (AR) has been increasingly studied in transportation, particularly for drivers and pedestrians interacting with automated vehicles (AVs). Previous research evaluated AR interfaces using online video-based questionnaires but lacked human-subject research in immersive environments. This study examined if prior online evaluations of nine AR interfaces could be replicated in an immersive virtual environment and if AR design effectiveness depends on pedestrian attention allocation. Thirty participants completed 120 trials in a CAVE-based simulator with yielding and non-yielding AVs, rating AR interface intuitiveness and crossing the road when safe. To emulate visual distraction, participants had to look into an attention-attractor circle that disappeared 1 second after the AR interface appeared. The results showed that intuitiveness ratings from the CAVE and previous online study correlated strongly ($r \approx 0.90$). Head-locked interfaces and familiar designs (augmented traffic lights, zebra crossing) yielded higher intuitiveness scores and quicker crossing initiations than vehicle-locked interfaces. Vehicle-locked interfaces were less effective when the attention-attractor was on the environment's opposite side, while head-locked interfaces were unaffected. In conclusion, this 'AR in VR' study showed strong congruence between intuitiveness ratings in a CAVE-based study and online research, and emphasizes the importance of interface placement in relation to user gaze direction.

Keywords: augmented reality, pedestrian-vehicle interaction, automated vehicles, CAVE

1. Introduction

Road accidents rank among the top ten causes of human fatalities in middle-income countries, as reported by the World Health Organisation (2020). Academia and industry have been researching a number of ways to curb this problem, such as through the use of smart infrastructure (Banks et al., 2018; Sewalkar and Seitz, 2019; Toh et al., 2020) and connectivity. The future urban environment will likely be more connected than today (Alam et al., 2017), and traffic partners may communicate between each other in order to facilitate traffic flow (Cao et al., 2022), solve right of way conflicts (Li et al., 2019), and enhance road safety for vulnerable road users, including pedestrians.

A specific solution is the use of external human-machine interfaces on automated vehicles (eHMIs), which could take the form of projections onto the road, LED strips, LED screens, anthropomorphic elements, amongst other examples (see Bazilinskyy et al., 2019; Dey, Habibovic, Pfleging, et al., 2020; Rouchitsas & Alm, 2019, for reviews of such interfaces). Although eHMIs have been shown to effectively convince VRUs whether or not to cross in front of approaching vehicles (De Clerq et al., 2019), there exist a number of drawbacks related to current designs, including the use of text, especially if the language is unfamiliar, cases where the eHMI needs to signal to a single pedestrian in a group, cultural differences in interpretability, and lack of standardisation across designs (De Winter & Dodou, 2022; Rasouli & Tsotsos, 2020; Tabone et al., 2021a; Tabone et al., 2023; Weber et al., 2019).

Augmented Reality (AR) has been proposed as a new type of communication in traffic, and as a possible solution to the aforementioned problems with eHMIs. In particular, AR offers the possibility of sending a customised signal to an individual pedestrian in a group (Tabone et al., 2021b; Tabone et al., 2023; Tran et al., 2022). Most AR studies so far have been conducted around the driver as the user (Calvi et al., 2020; Colley et al., 2021; Currano et al., 2021; Kim et al., 2018; Mukhopadhyay et al., 2023; Pichen et al., 2020), or as a navigation assistant to VRUs (Bhorkar, 2017; Dancu et al., 2015; Dong et al., 2021; Ginters, 2019; Tran et al., 2023). However, more recently, studies have explored the use of AR as an assistive technology for pedestrians, to guide them in making safe crossing decisions by including road projections of zebra crossings, arrows, and safe paths (Hesenius et al., 2018; Li et al., 2022; Pratticò et al., 2021; Tran et al., 2022; Tran et al., 2023), visualisation of obstructed vehicles, or the collision time and conflict points (Matviienko et al., 2022; Tong & Jia, 2019; Von Sawitzky et al., 2020), and car overlays (Tran et al., 2022).

Nine novel AR interfaces for pedestrian-AV interaction were previously developed in Tabone et al. (2021b) to support pedestrians in crossing an urban road. These interfaces were designed using an experience-based approach through a theoretically-informed brainstorming session and based on expert perspectives extracted from Tabone et al. (2021a). An expert ('genius')-based design method was employed (Saffer, 2010), where the designers drew upon known set theories such as predictive aiding, ecological interface design (Kardar & Shaw 2000), redundancy gain, and the proximity compatibility principle, amongst others (Wickens et al., 2004). Each AR interface was designed with two states (non-yielding, and yielding) to represent whether a vehicle would stop or not to that pedestrian, since the goal of these interfaces are to assist pedestrians in the decision to cross the road in front of an approaching AV.

Tabone et al. (2023) assessed these AR interfaces through an online questionnaire completed by 992 respondents in Germany, the Netherlands, Norway, Sweden, and the United Kingdom. The nine interfaces were recorded in a virtual reality (VR) environment and presented as videos to the respondents, who rated the interfaces for their intuitiveness and convincingness, aesthetics, and usefulness. Moreover, respondents were asked to provide free-text comments to further support their choices. Results indicated preference towards interfaces which employed traditional and familiar design elements from existing traffic, as well as head-up displays (HUDs). These insights were possible through the statistical and qualitative thematic analyses, which also revealed a number of unintended effects of certain designs.

Despite the rich body of information that was extracted from the online questionnaire study, one of the limitations of Tabone et al. (2023), was that it did not offer high ecological validity and presented only low perceived risks to the participants. A possible solution to this problem would be to use a VR simulation method that embodies the participant. The use of an immersive environment is important to test the distributed attention of participants, since in real traffic, accidents do occur because, similar to drivers, pedestrians may fail to look at the right object, at the right time (Lanzer et al., 2023; Lee, 2008; Ralph, & Girardeau, 2020).

Two possible VR simulation methods are to use a head-mounted display (HMD) or a CAVE (CAVE automatic virtual environment). An advantage of a CAVE setup is that it allows participants to see their bodies as they move around (Cordeil et al., 2017). Previous experiments with pedestrians in a CAVE investigated pedestrians' overreliance on AVs equipped with eHMIs (Kaleefathullah et al., 2022) and crossing behaviour of pedestrians on a road with continuous traffic (Kalantari et al., 2023; Mollaro et al., 2016). In this study, we extend the research of Tabone et al. (2023) by assessing the nine AR interfaces delineated in Tabone et al. (2021b) within a CAVE-based pedestrian simulator, with the objective to examine the generalizability of findings from online questionnaire studies.

There exist fundamental differences among various types of Augmented Reality (AR) designs (Arena et al., 2022; Carmigniani et al., 2011). Some AR designs are head-locked, wherein the message displayed follows the gaze of the participant. This feature ensures that the message is always visible to the pedestrian, allowing the user to benefit regardless of where the pedestrian is looking (Tabone et al., 2021a). Alternatively, AR systems may be positioned on the road infrastructure (Hesenius et al., 2018). The supposed advantage of this approach is that the interface can be found at known and expected locations (e.g., on the other side of the road, on the road surface). However, a disadvantage of this approach is that the AR interface may be overlooked when the participant glances left or right before deciding to cross the road. Finally, AR interfaces may be locked to the vehicle, just like eHMIs are. This approach has the advantage that the AR interface is congruent with the vehicle's motion, such that the pedestrian can process the implicit communication of the vehicle concurrently with the explicit AR signal. Furthermore, this approach eliminates the need for participants to distribute their attention towards multiple elements. However, a possible downside of vehicle-locked interfaces is that the vehicle must be identified before the pedestrian can benefit from the AR interface. For example, if a vehicle approaches from the right while the participant happens to be looking to the left and does not immediately scan to the right, identifying the AR interface may be inefficient.

In this study, we study effects of initial visual attention, using a novel technique to guide the pedestrian's initial attention towards specific regions of the road before the arrival of the vehicle. Our approach involved simulating the behaviour of pedestrians who initially fail to observe the approaching vehicle, as attention resources are finite (e.g., Wickens et al., 2004; Ralph & Girardeau, 2020), and individuals cannot attend to the entire traffic scenario simultaneously. Specifically, we employed an attention-attractor circle in the form of circles placed either on the left, front, or right side of the scene to investigate our hypothesis about whether the effectiveness of the AR interface type would interact with the participant's initial attentional position. Specifically, the attention attractor was to be looked at for 1 s before any interface would appear in the environment. The aforementioned method was used to emulate and enforce initial distraction towards a certain region of the environment.

The aim of the experiment is to examine the effects of the nine different AR interfaces (previously described in Tabone et al., 2021b; Tabone et al., 2023) on pedestrian crossing behaviour and perceived intuitiveness, relative to a no-AR baseline condition, as well as their relative effects on each other. It is expected that the quality ratings observed in Tabone et al. (2023a) will be replicated in the present immersive CAVE-based environment. Moreover, the nine AR interfaces are anticipated to be regarded as clearer than the no-AR baseline conditions.

Additionally, the current study investigated the effect of the above-mentioned position-based attention attractor on pedestrian crossing behaviour and perceived intuitiveness for different AR interfaces. It is hypothesised that there will be an AR/attention-attractor position interaction, where head-locked AR interfaces (e.g., *Nudge HUD*, *Pedestrian lights HUD*) will be perceived as more intuitive when the attention-attractor is presented on the left relative to the otherwise mapped interfaces, while infrastructure-locked (road-mapped) AR interfaces (e.g., *Augmented zebra crossing*, *Fixed pedestrian lights, Virtual fence*) will be perceived as more intuitive when the attention-attractor is presented in the centre location of the CAVE. Finally, vehicle-locked AR interfaces (e.g., *Planes on vehicle*, *Conspicuous and looming planes*, *Field of safe travel*, *Phantom car*) are expected to be perceived as more intuitive when the attention-attractor circle is presented to the right.

2. Methods

2.1 Participants and Recruitment

Thirty participants (20 identified as male, 9 identified as female, and 1 was unspecified), aged between 22 and 53 (M = 31.50, SD = 7.98) were recruited for the study. The 30 participants were of 12 different nationalities, namely British (10), Chinese (5), Greek (3), Indonesian (3), Malay (2), German (1), Maltese (1), Norwegian (1), Romanian (1), Saudi (1), Turkish (1), and Zimbabwean (1). From the participant pool, 56.7% (*n* = 17) indicated that they had never been in a CAVE, while

43.3% (*n* = 13) indicated that they had. Further general characteristics of the 30 participants, which were collected in the demographic survey, are available in the Appendix A2.

Participants were recruited using an opportunistic sampling approach. Internal emails were sent to a pool of people interested in participating in the University of Leeds Driving and Pedestrian Simulation studies, and to various schools at the University of Leeds. Moreover, adverts were posted on a Facebook group consisting of students studying at the same university, and a further group composed of residents of Leeds.

Criteria for participating in the study were highlighted in both the emails and adverts. Specifically, participants were only eligible to participate if they were over the age of 18, had a good command of English, did not suffer from severe mobility issues, and did not suffer from epilepsy, claustrophobia, or feelings of disorientation. Moreover, we asked participants to wear lenses if they had prescription glasses, and if they had long hair, to tie it back on the day of the experiment due to the eye tracking equipment.

Participants were able to select an available time slot using an online calendar system which was linked to the experimenter's and simulator's calendars. The experiment ran for two weeks between June and July 2022, and the participants were financially compensated with a £15 Amazon gift voucher for their time spent, which was roughly 60–90 minutes. The experiment was approved by the University of Leeds Research Ethics Committee under ethics reference number LLTRAN-150.

2.2 Apparatus

The experiment was created using Unity v.2020.3.35f1 (Unity, 2022) and performed in the Highly Immersive Kinematic Experimental Research (HIKER) simulator located at the University of Leeds (University of Leeds, 2022). The HIKER is a 9 × 4 m Cave Automatic Virtual Environment (CAVE) simulator, composed of eight 4K high resolution (3840 × 2400 px) projectors, and 10 Vicon Vero 2.2 IR cameras, which were calibrated and controlled using Vicon Tracker 3.9. The simulator supports both stereo and mono modes. For this experiment, mono mode was used.

Gaze data were sampled at 50 Hz and collected using the Tobii Pro Glasses 2 (firmware 1.25.6-citronkola-0; head unit 0.062) mobile eye-tracker, which was operated and calibrated using the Tobii Controller Software v.1.114.20033. Participants' verbal statements were recorded using an Olympus VP-20 microphone. A Logitech web camera mounted on a tripod was used to record the entire experiment per participant in low resolution. High-resolution video clips for dissemination were recorded using a GoPro Hero 10 camera and an iPhone 13 Pro.

The entire experiment was run on an eight-computer rack, with seven Image Generator (IG) machines brandishing an Intel® Core™i9-7900X CPU @ 3.30 GHz, 32 GB RAM, and 8 GB Nvidia Quadro P6000, and the host machine equipped with Intel® Core™i9-7900X CPU @ 3.30 GHz, 128 GB RAM, and 8 GB Nvidia Quadro P4000.

2.3 Augmented-Reality Interfaces

The interfaces that were evaluated in this simulator study were adopted from the designs in Tabone et al. (2021b), and identical to the VR implementation used in Tabone et al. (2023a). In total, nine AR interfaces were designed and developed as functioning AR prototypes (Tabone et al., 2021b).

In general, the interfaces were split into three categories: interfaces that were mapped to the road, to the vehicle, and the user's head position (ie. HUDs). There are three interfaces that were mapped to the road:

- *Augmented zebra crossing* (Labelled 1 in Figure 1), which is a conventional traditional zebra crossing
- *Fixed pedestrian traffic* lights (5), which depicts a familiar pedestrian traffic light design across the road, and a
- *Virtual fence* (6), which displays semi-translucent walls around a zebra-crossing and a gate that opens in the yielding state.

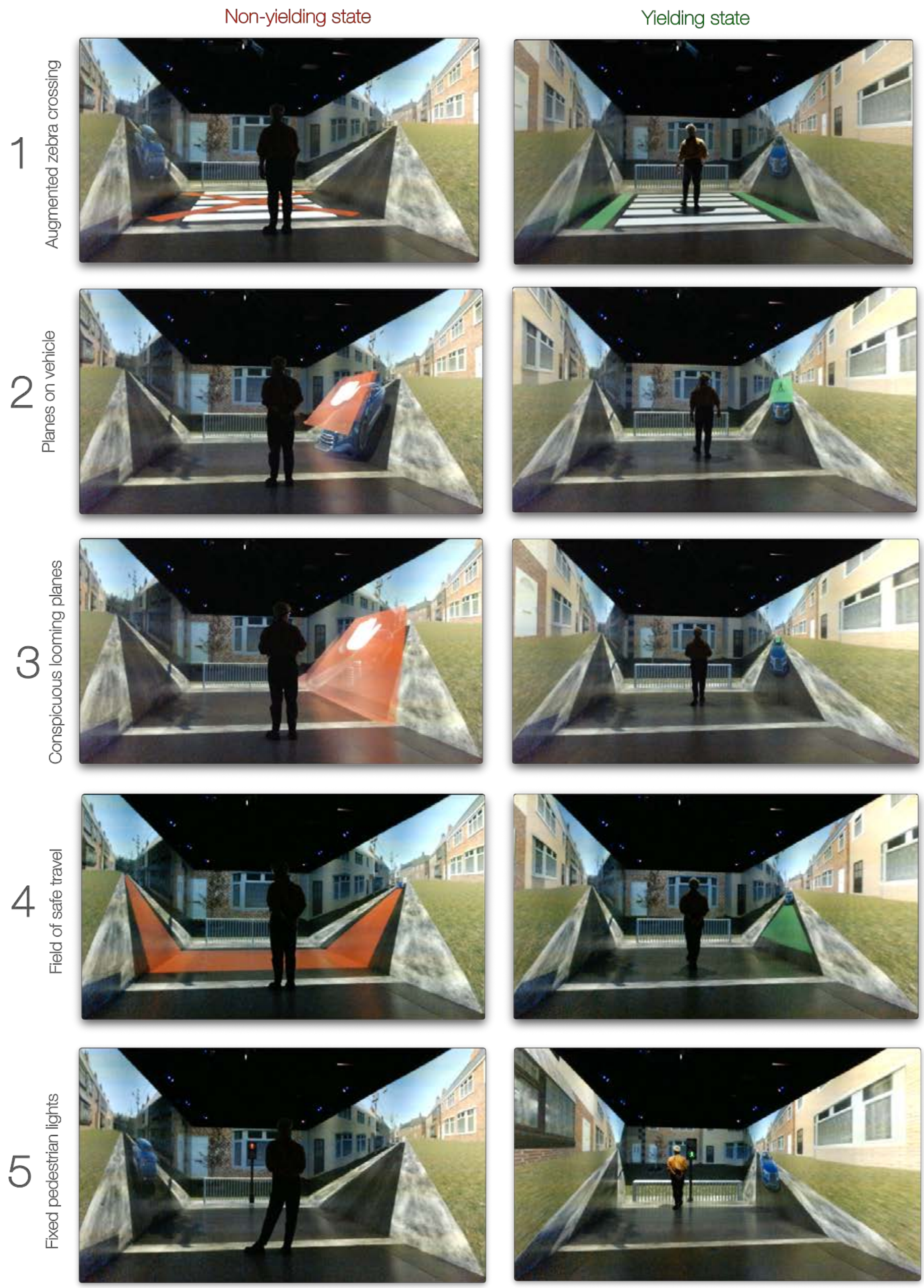
The interfaces that were mapped to the vehicle include:

- *Planes on the vehicle* (2), which displays a plane on the windshield area of the vehicle,
- the *Conspicuous looming planes* (3), which is a scaling version of (2), as it grows or shrinks as the vehicle approaches the pedestrian depending on the AV's yielding state,
- the *Field of safe travel* (4) which projects a field on the road in front of the vehicle to communicate safety, and
- the *Phantom car* (7), which projects the vehicle's predicted future motion.

The final category of interfaces are head-up displays that are head-locked to the user's head position, i.e., they follow and remain in the user's field of view. These are:

- the *Nudge HUD* (8), which displays text and icons, and
- the *Pedestrian lights HUD* (9), which displays a head-locked version of the pedestrian traffic lights.

Interfaces mapped to the road or the vehicle were positioned within the environment or attached to the vehicle, while the HUDs moved with the participant's camera view. The *Nudge HUD* measured 65 cm in width, and 20 cm in height and was situated 2.5 m away from the participant, while the *Pedestrian lights HUD* measured 20 cm in width and 40 cm in height. The bottom edge of the HUDs aligned with the participant's eye level. In comparison to the online questionnaire study, the dimensions of the HUDs were reduced due to the restricted field of view (FOV) in the CAVE. The red (RGB: 244, 0, 0) and green (RGB: 32, 244, 0) colours selected for the non-yielding and yielding states were chosen based on their high intuitiveness score ratings for signalling 'please (do not) cross' (Bazilinskyy et al., 2020).


Non-yielding state
Yielding state
1 Augmented zebra crossing
2 Planes on vehicle
3 Conspicuous looming planes
4 Field of safe travel
5 Fixed pedestrian lights

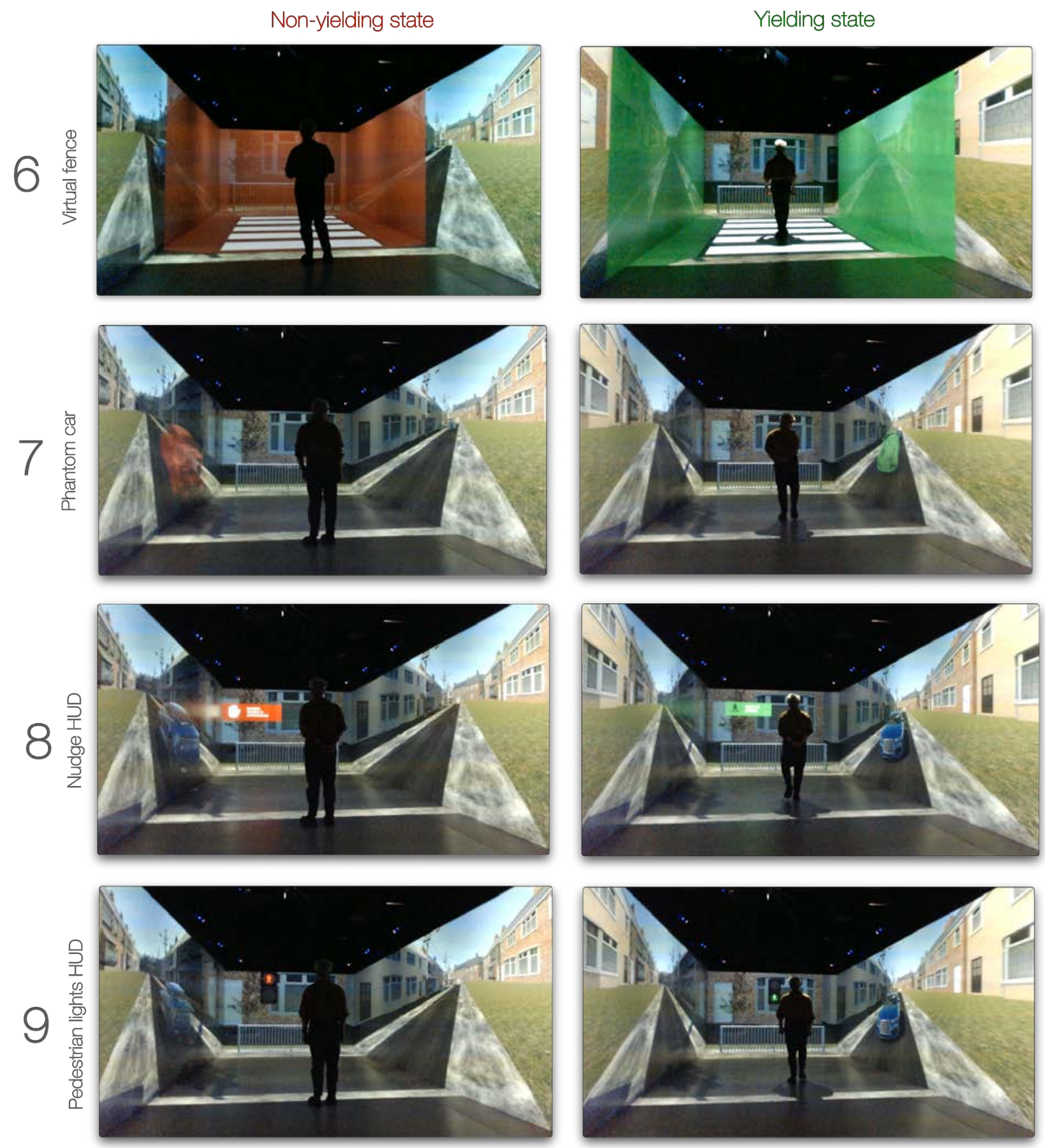


Figure 1. The nine AR concepts presented in the HIKER environment. Interfaces 1, 5, and 6 are projected on the road surface, 4 and 7 are also projected on the road surface, but mapped to the vehicle, while Interfaces 2 and 3 are projected on the vehicle. Interfaces 8 and 9 are head-locked.

2.4 Scenario Design

The current study adopted the road environment and vehicle behaviours from Kaleefathullah et al. (2022) and Tabone et al. (2023). The AV spawned out of sight (Figure 2, point A) and moved at a constant speed of 48 km/h (30 mph). The AR interfaces were activated when the AV was 43 m away in virtual space (Figure 2, point B) from the participant, who was located at point E. For non-yielding states, the vehicle maintained its initial speed of 48 km/h throughout. In contrast, for yielding states, the vehicle started decelerating at Point C, which was 33 m from Point E (the participant), at a rate of 2.99 m/s^2, and came to a complete stop at Point D, located 3 m from Point E.

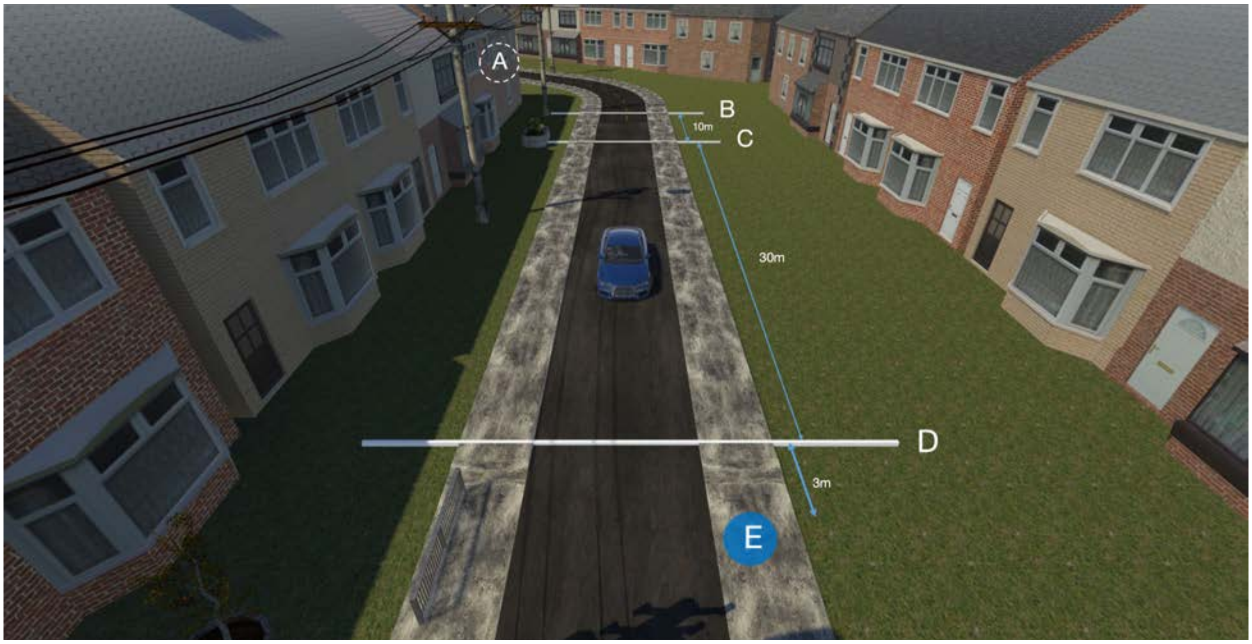


Figure 2. The simulation environment used for the HIKER experiment. Each salient point is demarcated by a label, together with the distance (in metres) between each point. A: spawn point, B: AR interface onset, C: AV deceleration onset, D: stopping point, E: participant location. Image adapted from Tabone et al. (2023).

In each trial, an attention-attractor circle was presented in the shape of a stationary open circle with cyan border (RGB: 0, 255, 255), which was presented in three different positions: left, front-centre (i.e., across the street), and to the right of the participant, at the start of each trial (Figure 3). Cyan was chosen as the colour since it has been suggested to be considered neutral, and without any meaning in current traffic (Bazilinskyy et al., 2020). The circle would disappear if the participant looked into the area enclosed by its circumference for 1 second. The detection of whether the participant was looking into the circle was conducted through the head-tracker. If this condition was met, the trial commenced.

Figure 3. A participant is seen standing on the blue marker and looking at the cyan attention-attractor circle, which is displayed in the centre position.

Two seconds later, the vehicle arrived at point B (see Figure 2) and the AR interface was activated (Figure 4iii). If the condition was an interface in yielding state, the vehicle started decelerating at point C (see Figure 2). The circle disappeared 1 s after the AR interface was activated (see Table 1 for the general timings of the trials). If this constraint was not not applied, participants may have looked at the car as soon as possible since they expect the car to appear from behind the corner of the road, or because they could see it in their peripheral vision.

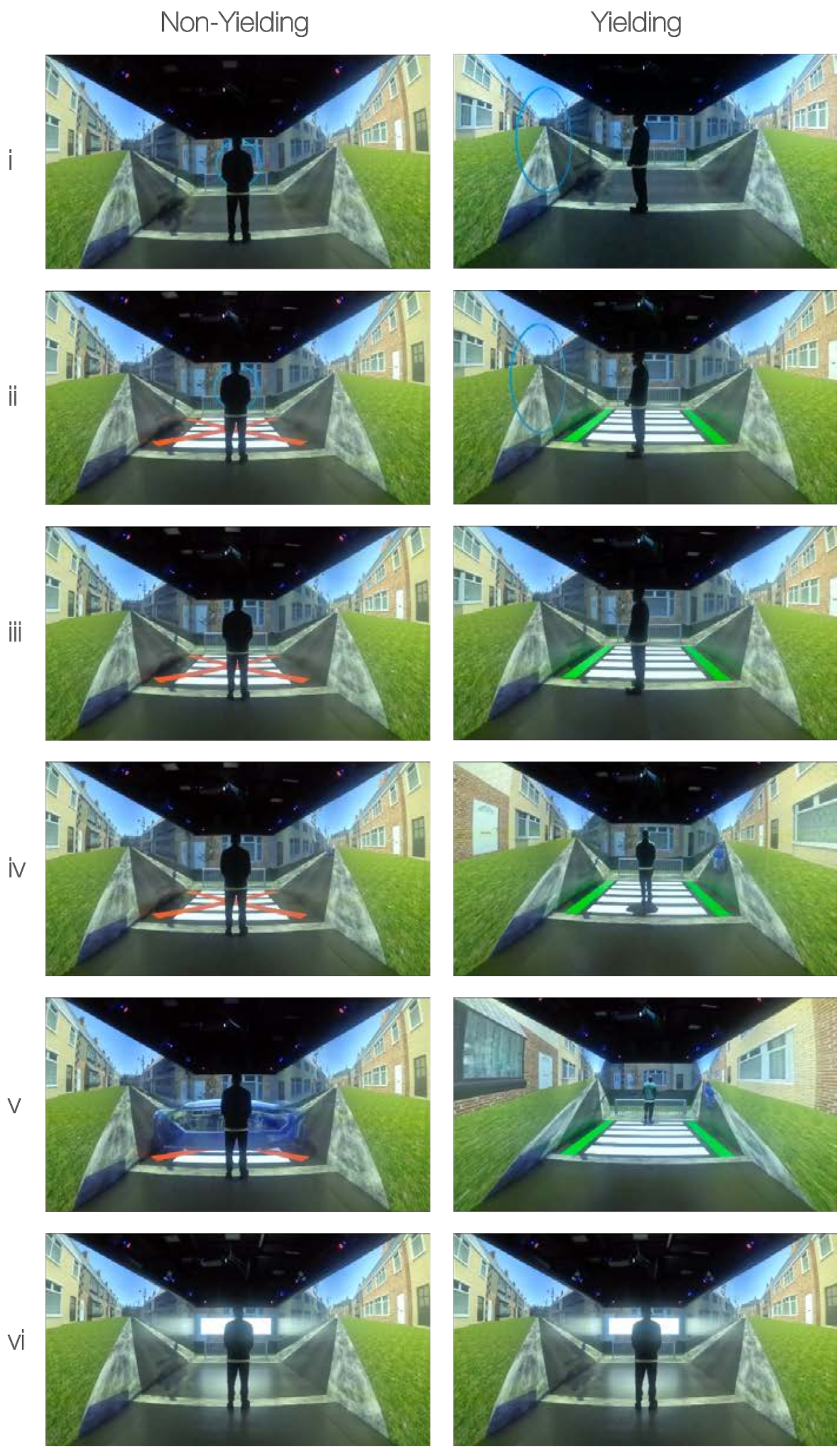


Figure 4. Timeline of the crossing experiment, with the left column representing the non-yielding condition, and the right column the yielding condition.

Table 1.
Timings of events of trials, centred around the moment the circle disappeared. The letters in parentheses correspond to the positions in Figure 2.

Event	**Elapsed time (s)**	
	Non-yielding condition	**Yielding condition**
Attention-attractor circle appears	-9.0	-9.0
Attention-attractor circle looked at for 1 s and vehicle starts driving (A)	-8.0	-8.0
Vehicle appears from behind curve	-3.0	-3.0
AR interface appears (B)	-1.0	-1.0
Vehicle starts to decelerate (C)	—	-0.2
Circle disappears	0.0	0.0
Vehicle comes to a halt (D)	—	3.8
Vehicle passes participant (E)	2.3	—
Intuitiveness question appears	7.0	8.8

2.5 Experimental Design

The experiment was a within-subject design, consisting of 120 trials per participant: 90 trials with a yielding AV and 30 trials with a non-yielding AV. More specifically, the yielding AV trials consisted of 10 AR interface conditions (i.e., including the Baseline conditions) x 3 circle positions x 3 repetitions, while the non-yielding AV trials comprised 10 AR interface conditions x 3 circle positions x 1 repetition. Each interface condition was presented in a block. Therefore, ten blocks were presented, and a counterbalancing technique was applied. Each block consisted of 12 trials (4 yielding/non-yielding conditions x 3 circle positions) which were presented in a randomised order.

After each trial, a question was projected on the centre screen of the CAVE. Participants had to verbally indicate their agreement from 1 (Strongly disagree) to 7 (Strongly agree), with the statement: "*The interface was intuitive for signalling: 'Please (do NOT) cross the road'*". In the case of the Baseline condition, which featured no AR interface, the word, 'interface' was replaced by 'situation'. Figure 5 demonstrates various moments in a trial.

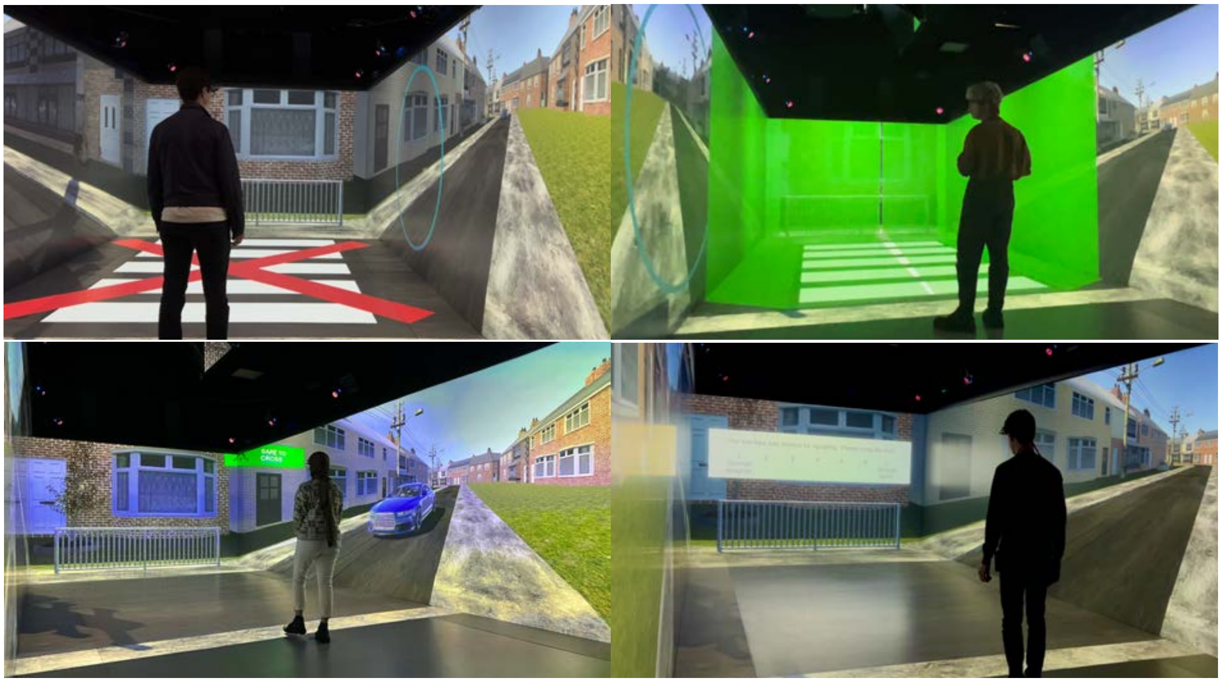


Figure 5. Top-left: The non-yielding state of *Augmented zebra crossin*g is projected just before the attention-attractor circle in the right position disappears. *Top-right:* Yielding state of *Virtual fence*, projected before the circle in the left position disappears. *Bottom-left: Nudge HUD* in a yielding state is seen projected in the CAVE. *Bottom-right:* The intuitiveness rating scale is displayed at the end of a trial.

2.6 Experimental Procedure

Participants were welcomed into the room containing the CAVE simulator by an experimenter. An information sheet containing the experiment protocol (also previously sent to the participant by email) was presented. If the participant agreed with all the information, they were asked to sign the consent form. Participants were reminded that they can choose to stop the experiment at any time. Participants were then asked to complete a demographics questionnaire, which was developed using the Qualtrics XM (Qualtrics, 2022) survey platform. Details of the questionnaire flow are presented in Appendix A1.

Following the demographics questionnaire, the participants were asked to wear overshoe covers to protect the HIKER flooring, and the Tobii eye tracker system, which was also equipped with infrared markers to track the head position and correctly align the CAVE to the participant's FOV. The eye tracker was subsequently calibrated, and the participant walked into the simulator and stood on a blue marker demarking the starting position for each trial (Figure 4i).

The experimenter then reminded the participant of some key points highlighted in the information sheet. They were asked to look into the cyan circle for 1 consecutive second (Figure 4ii). If the participant violated this rule, a beeping sound was automatically produced to draw the attention of the participant back to the circle. If the participant adhered to the instructions, the vehicle would come out from behind the corner.

Furthermore, participants were reminded that they might encounter various suggestive interfaces, and that they were to cross the entire road (from one curb to the other) only if they felt that it was safe to do so (Figure 4iv, right). If the vehicle drove past them (Figure 4v left), then they no longer needed to cross. Once they had crossed the road, and the intuitiveness question was displayed (Figure 4vi), they were allowed to walk back to the starting position and read out their answer there. The participants were also reminded about a digital mesh that appears across the nearest CAVE wall if they get too close, to warn participants to stop walking and avoid crashing into a physical barrier.

For improved understanding, two practice trials of the baseline conditions, one with a non-yielding vehicle and one with a yielding vehicle, were conducted prior to the initiation of the main experiment. When the participant indicated that they had understood the procedure, the experimenter announced that the actual experiment was about to begin. At that point, the participant was asked to return to the blue marker, and the experimenter started the experiment.

Participants then completed 120 trials. After each block, a 3-minute interview was recorded with the participant. The interview typically began by asking the participant whether they were comfortable. If the participant mentioned any form of discomfort, then the MISC scale was administered. Next, they were asked "*what did you think of this particular interface/situation?*", and the participant was prompted to elaborate further on their answers. The line of questioning about the interface continued by asking the participant about preference between the red and green states, and whether their crossing decisions were based on the interface. Participants were corrected if they began commenting about the VR environment rather than the interfaces themselves.

Following all the trials, the participant was invited back to the table where they had signed the consent form. There they were presented with a sheet containing a table with screenshots of all the nine interfaces in both their yielding and non-yielding states side by side. The participant was asked to assign a rank from 1 to 9 next to each interface according to their preference, with 1 being the most preferred, and 9 being the least preferred. Each number could only be assigned once. When the participant had finished with the ranking, they were thanked and rewarded for their time.

2.7 Data Logging

The vehicle's position and speed were logged during the experiment at a frequency of 120 Hz, to produce a total of 3600 log files. For the gaze data analysis (logged at 50 Hz) and head-tracking data (logged at 120 Hz) the VR environment was segmented into a number of areas of interest:

- **Road1, Road2, Road3**: We used Road1 (near distance), Road2 (medium distance), and Road3 (far distance) segments (see Figure 6) to explore the distribution of the participants' attention as the AV approaches. Given that the AV approached from the right in all trials, no areas of interest to the left were created.
- **Road**: the rest of the road.

- **Car**: main body of the vehicle (Figure 6, inset).
- **Circle**: the region encompassed by the cyan attention-attractor circle.
- **AR Concept**: the region where the AR interface is projected.
- **HUDReg**: the region where the HUD interfaces (8 and 9) are projected. The HUD region follows the participant's head rotation and is always in front of the participant, 2.5 m ahead, and 0.25 m above the participant's head.
- **TrafficLightReg**: the region where the Fixed Traffic LIghts would be projected, i.e., a static region in front of the pedestrian.
- **WindScrReg**: the windscreen region on the vehicle (Figure 6, inset).
- **Other**: any other non-segmented area in the CAVE, or outside of it.

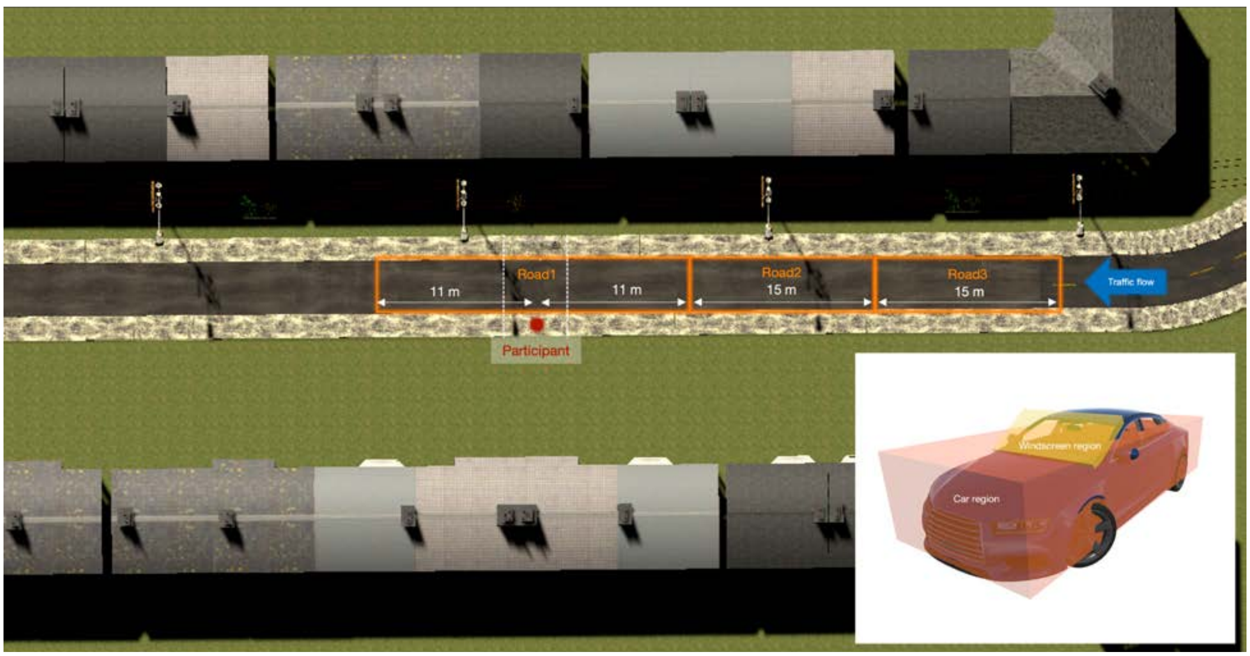


Figure 6. Top-down view of the environment. The road was segmented into three different regions for logging purposes. Inset: Two collider regions labelled as 'Car' and 'WindScrReg' were superimposed over the car 3D model in the environment. The labels were logged when the head-pose and eye-tracker vectors intersected with the colliders.

These regions were active for all tested conditions, except the *AR Concept* region for the Baseline condition. Therefore, some segmented regions were prioritised over others when a gaze or head vector intersected in a region where multiple segmented layers may have been overlaid on one another. Priorities were as follows: The *AR Concept* region took precedence over all other regions, followed by the *Car*, *WindScrReg*, *TrafficLightReg*, *HUDReg, Road1, Road2*, *Road 3* ,*Circle* , *Road, and Other,* in the order presented.

2.8 Processing of Post-Block Interviews

The post-block interviews were analysed using a novel approach, outlined by Tabone and De Winter (2023). Specifically, the 300 post-block voice recordings were transcribed using Otter.ai (2023), an AI tool which offers an audio to text transcription service. The transcripts were exported as text files where the text was automatically split by the different speakers, and timestamps by Otter.ai. Each transcript was then automatically submitted to OpenAI's ChatGPT-4 API (version: March 14, 2023). The API's temperature setting, which controls the level of the output's randomness, was set to 0.

The following prompt was used to summarise each transcript file, encouraging ChatGPT to extract multiple strengths and weaknesses from the interview: "*Based on the participant's responses in this interview, reports three strengths and three weaknesses about the AR interface. Start the strengths with "STRENGTHS:" and start the weaknesses with "WEAKNESSES:*". Once all transcripts were summarised, the output summaries per interface (i.e., all summaries of an interface, combined) were submitted together to the ChatGPT API and summarised once again using the following prompt: "*Based on all the summaries below, summarise the strengths and weaknesses of the interface concisely. Do not overly mention general characteristics such as ''the interface is intuitive'', but report specific aspects that could aid designers. Keep it short, specific, and interesting, with a maximum of 4 sentences in total. Start the strengths with "STRENGTHS:" and start the weaknesses with "WEAKNESSES:"*.

2.9 Dependent Measures

The following dependent measures were calculated:

- ***First glance at the AR interface***: this measure was calculated using the Tobii eye-tracker data. Ray tracing was applied to determine at first moment the gaze vector intersected with the AR interface. This measure was not available for the Baseline condition.
- ***Participant crossing initiation time***: the measure was computed for each trial in which the AV yielded. The crossing initiation time was calculated by subtracting the moment the participant's position in the CAVE environment exceeded a set threshold (corresponding to the edge of the road) from the moment at which the attention-attractor circle disappeared.
- ***Intuitiveness ratings***: the self-reported intuitiveness rating was available for each trial.
- ***Interface ranking***: the mean rank, per participant per AR interface was calculated from the responses to the post-experiment ranking questionnaire. In addition, the distribution of the ranking was also analysed from a produced matrix (Appendix A3).
- **Sentiment score**: the post-block questionnaire transcript files were also analysed to generate a sentiment score, using the same ChatGPT prompt which analysed the sentiment for the online questionnaire respondent open-question responses (Tabone & De Winter, 2023): "*Looking at the participant's responses, score the interface, from 1 to 100. Only report a number between 1 and 100, rounded to two decimals*". Note that the prompt provides no dimension, such as 'intuitiveness', an approach which allows GPT-4 to provide a generic sentiment rating (Tabone & De Winter, 2023).

In addition to calculating the above dependent measures, we plotted a number of graphs to better understand how participants used the AR interfaces. In order to visualise where the participants looked at during the trials, all timings were centred and normalised around the moment that the attention-attractor circle disappeared. Hence, all time values were rounded to the nearest multiple of the sampling rate (120 Hz). Then, for each 0.01 s timestamp between -

9.0 s and 9.0 s, the counts for each time a gaze region was gazed at (eye-tracking) were registered. These gaze counts against time were converted to percentages (%) and plotted against time (s) for a number of conditions (such as the yielding state, and the circle position). The produced plots would provide a visualisation of where participants distributed their gaze across areas of interest (AOI) in the CAVE.

2.10 Statistical Analysis

In order to judge the similarity of the simulator results with the obtained results from the online questionnaire study, the Pearson product-moment correlations of means for the AR interfaces were calculated. Moreover, to understand the effect of the attention-attractor circle on the dependent variables, a two-way repeated measures ANOVA of crossing initiation times and the time of the first glance towards the AR interface was conducted, with the AR interface and circle position as within-subject factors. Within-subject confidence intervals of the means were also generated according to a method presented by Morey (2008). Pairwise comparisons between the three circle positions were conducted using paired t-tests. To account for multiple comparisons, a Bonferroni correction was employed, resulting in an adjusted alpha value of 0.05 / 3 (given that there were three pairs for comparison).

3. Results

All 3600 trials were completed successfully. There were only three instances where blocks had to be restarted because of technical disruptions (e.g., the eye-tracker switched off). However, in each case, the previous intuitiveness scores were retained, and any missing log data re-recorded. Hence, there was no missing data by the end of the experiment.

3.1 Objective 1: Replication of Intuitiveness Ratings: CAVE study vs. Online Study

Figure 7 shows scatter plots reporting the mean intuitiveness scores of each interface for both the CAVE simulator experiment, and the online questionnaire study (Tabone et al., 2023). Since the online study did not have a Baseline condition, its mean intuitiveness score (yielding vehicle: 4.80 [*SD* = 1.37], non-yielding vehicle: 5.48 [*SD* = 1.53]) was omitted from the plot.

The mean intuitiveness score from the HIKER study correlated strongly with the mean intuitiveness score from the online study ($r = 0.91$ for yielding vehicles, and $r = 0.90$ for non-yielding vehicles, respectively, $n = 9$). Similar to the online questionnaire study, the *Nudge HUD*, *Fixed pedestrian lights, Augmented zebra crossing, Pedestrian lights HUD, and Virtual fence* interfaces scored the highest, while the *Phantom car* yielded the lowest intuitiveness rating.

Table 2. *Correlation matrix for measures relating to the online questionnaire study (Tabone et al., 2023), and the simulator study conducted in the HIKER*

	1	2	3	4	5	6
1. Hiker: Intuitiveness (non-yielding vehicle)						
2. Hiker: Intuitiveness (yielding vehicle)	0.73					

3. Hiker: ChatGPT sentiment score	0.85	0.82				
4. Hiker: Mean preference rank	-0.77	-0.90	-0.94			
5. Hiker: Crossing initiation time	-0.51	-0.87	-0.59	0.63		
6. Online: Intuitiveness (non-yielding vehicle)	0.90*	0.83	0.86	-0.71	-0.77	
7. Online: Intuitiveness (yielding vehicle)	0.83	0.91*	0.90	-0.79	-0.90	0.94

Note. n = 10 for measures 1, 2, 3, and 5. n = 9 for measures 4, 6, and 7.
* These two correlation coefficients are depicted in scatter plots in Figures 7 and 8.

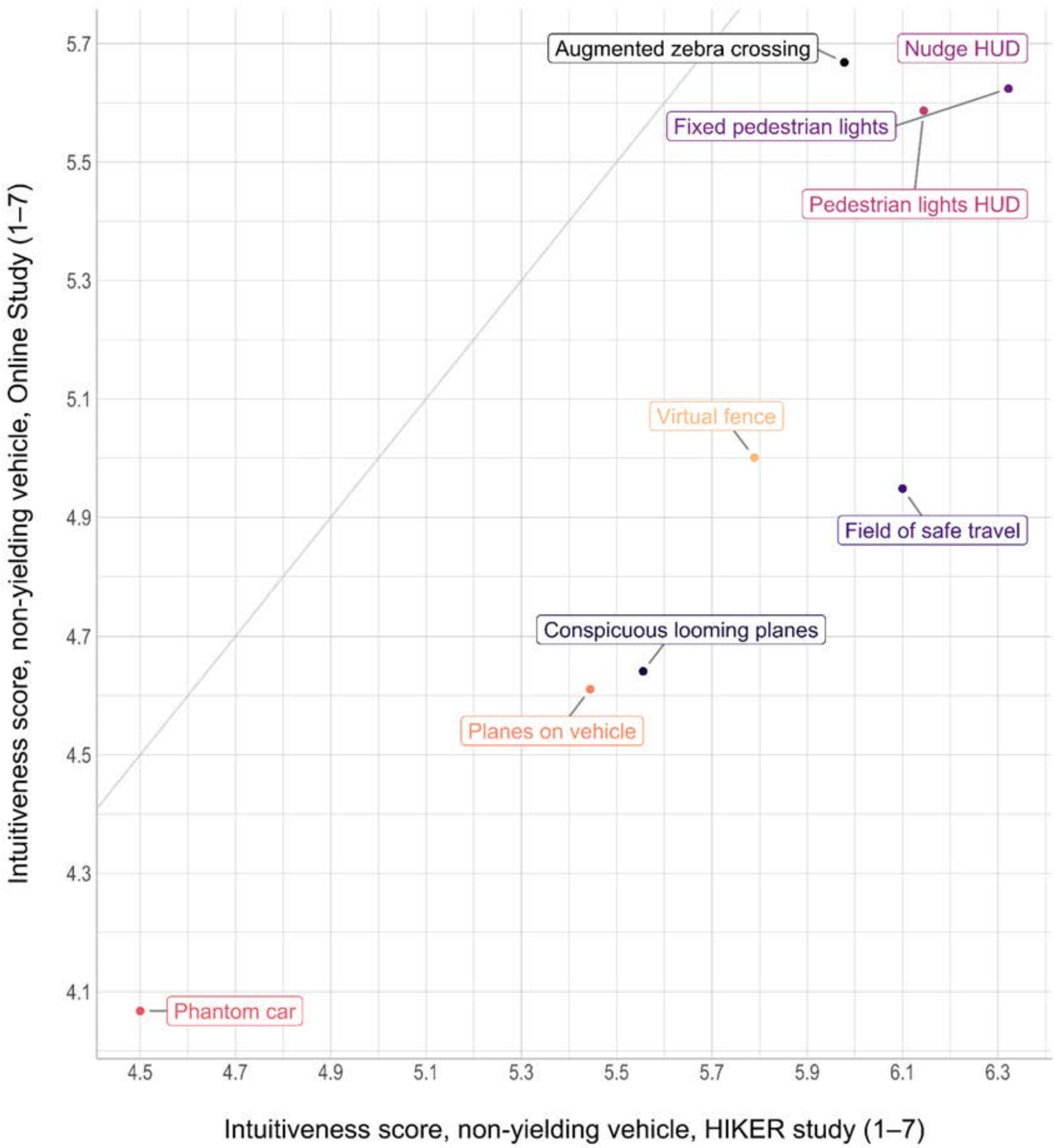


Figure 7. Scatter plot of the mean intuitiveness rating per AR interface in their non-yielding state, as reported in the current HIKER experiment and the previous online questionnaire study (Tabone et al., 2023).

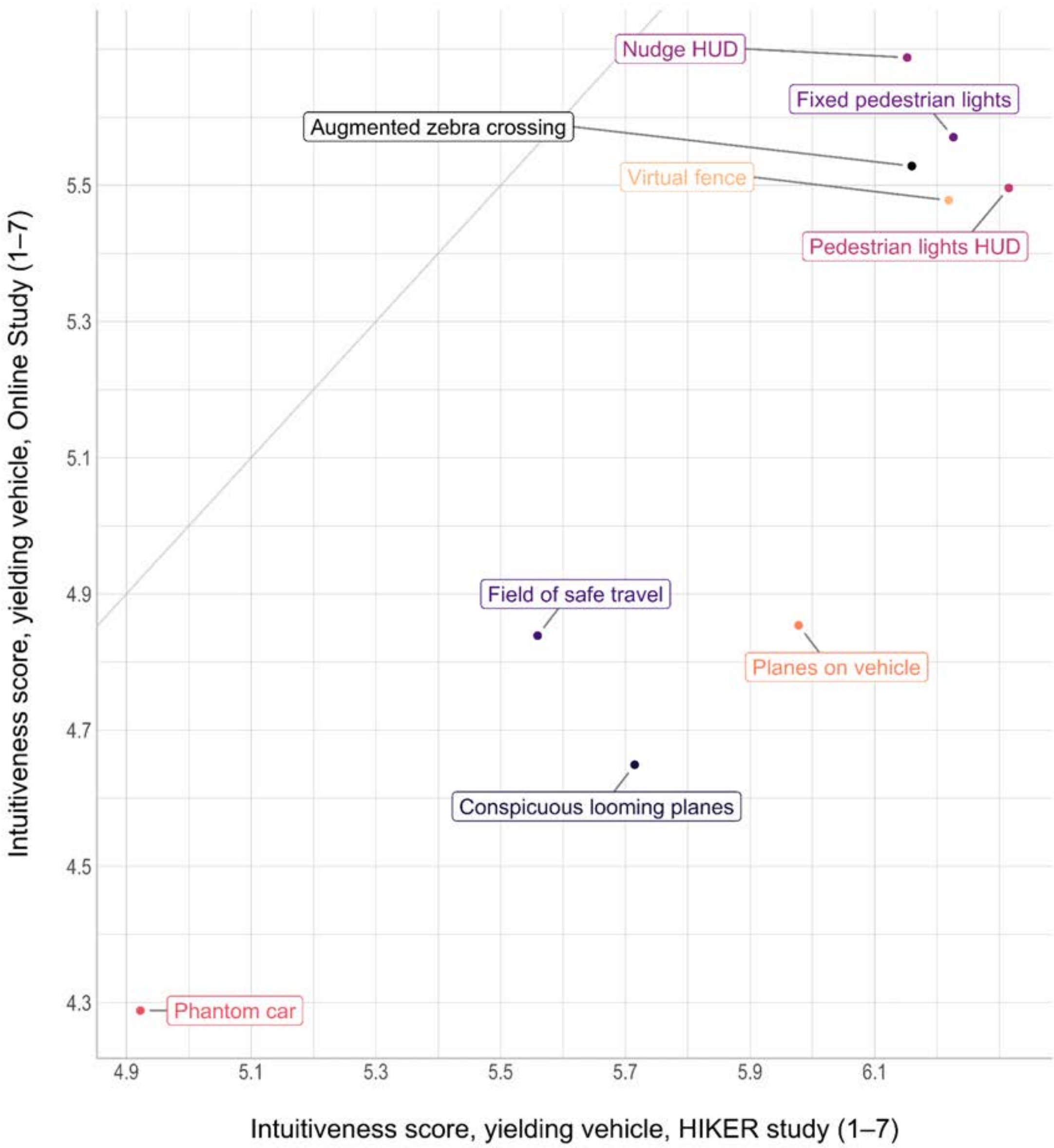


Figure 8. Scatter plot of the mean intuitiveness rating per AR interface in their yielding state, as reported in the current HIKER experiment and the previous online questionnaire study (Tabone et al., 2023).

3.2. Objective 2: AR interface/Attention-Attractor Position Interaction

3.2.1 Crossing Initiation Time

According to a repeated-measures ANOVA of the crossing initiation time, with interface condition and circle position as within-subject factors, there was a significant effect of interface condition,

$F(9, 252) = 17.98$, $p < 0.001$, partial $\eta^2 = 0.39$, a significant effect of circle position, $F(2, 56) = 56.163$, $p < 0.001$, partial $\eta^2 = 0.67$, and a significant interface condition × circle position interaction, $F(18, 504) = 2.13$, $p < 0.001$, partial $\eta^2 = 0.22$.

Figure 9 shows the means and 95% confidence intervals of the crossing initiation times for all AR interfaces. Among the road-mapped interfaces, the *Fixed pedestrian lights* (which were always positioned on the opposite side of the street) demonstrated optimal performance when the attention-attractor circle was also presented centrally. Regarding the *Augmented zebra crossing*, the central and right circle presentation positions exhibited faster crossing initiation times as opposed to the left presentation, while the *Virtual Fence* performance remained relatively consistent irrespective of the circle presentation position.

Concerning the vehicle-mapped AR interfaces, a consistent trend emerged, wherein superior performance was observed when the circle was presented on the right—the direction from which the AV approached.

Lastly, with respect to the HUDs and Baseline condition, a discernible pattern emerged, where the left circle presentation was less advantageous compared to central or right presentations. However, this effect was not statistically significant for the *Baseline* condition and *Pedestrian lights HUD* and was considerably smaller in magnitude than the vehicle-mapped interfaces.

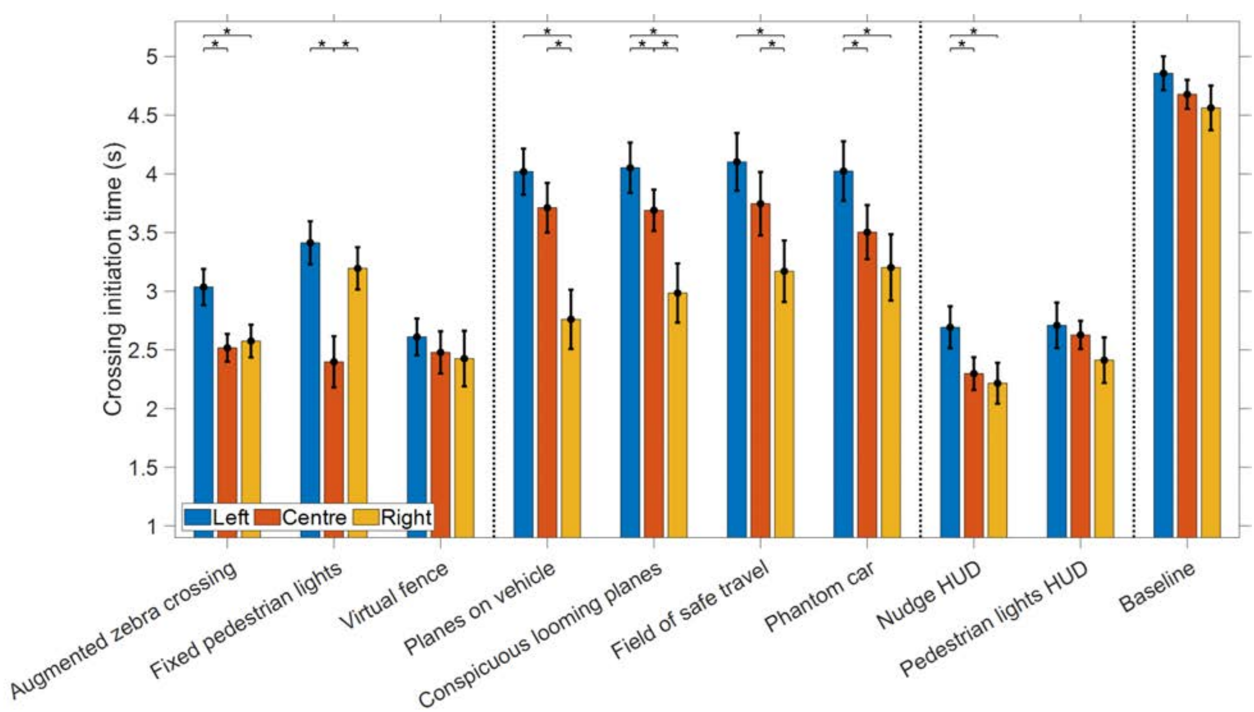


Figure 9. Bar plot of the mean crossing initiation time as affected by the attraction-attractor position, for each AR interface. Vertical lines delineate the road-mapped interfaces, vehicle-mapped infaces, HUDs, and baseline condition.

3.2.2 Time of First Glance at AR interface

According to a repeated-measures ANOVA of the time of the first glance at the AR interface, with interface condition, and circle position as within-subject factors, showed a significant effect of interface condition, $F(8, 200) = 21.23$, $p < 0.001$, partial $\eta^2 = 0.46$, a significant effect of circle position, $F(2, 50) = 36.58$, $p < 0.001$, partial $\eta^2 = 0.59$, and a significant interface condition × circle position interaction, $F(16, 400) = 15.30$, $p < 0.001$, partial $\eta^2 = 0.38$.

As a manipulation check for the experiment, the results presented in Figure 10 generally align with the crossing initiation times shown in Figure 9. Specifically, when the attention-attractor circle was centrally positioned, the *Fixed pedestrian lights* were the first to be glanced at, while the vehicle-mapped interfaces drew attention more quickly when the circle was displayed on the right as opposed to the centre or left. Regardless of the circle position, the HUDs were promptly noticed. Furthermore, the *Virtual fence* quickly captured attention, which can likely be attributed to its significant size.

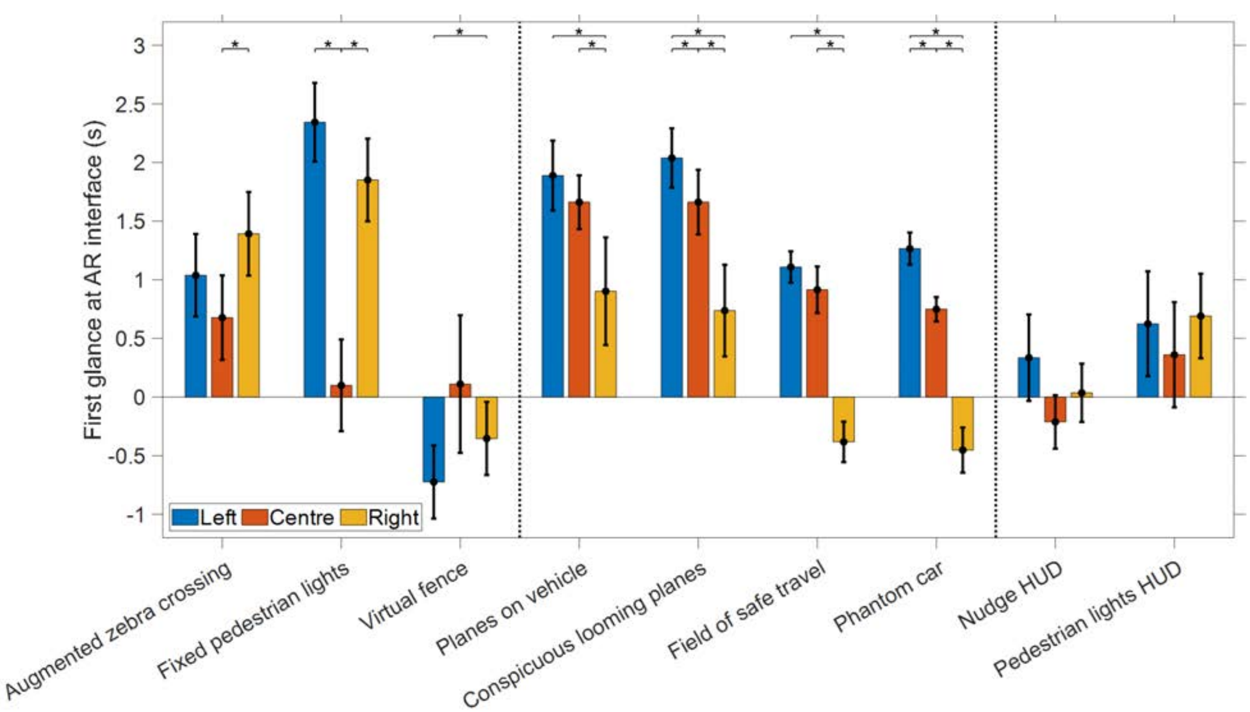


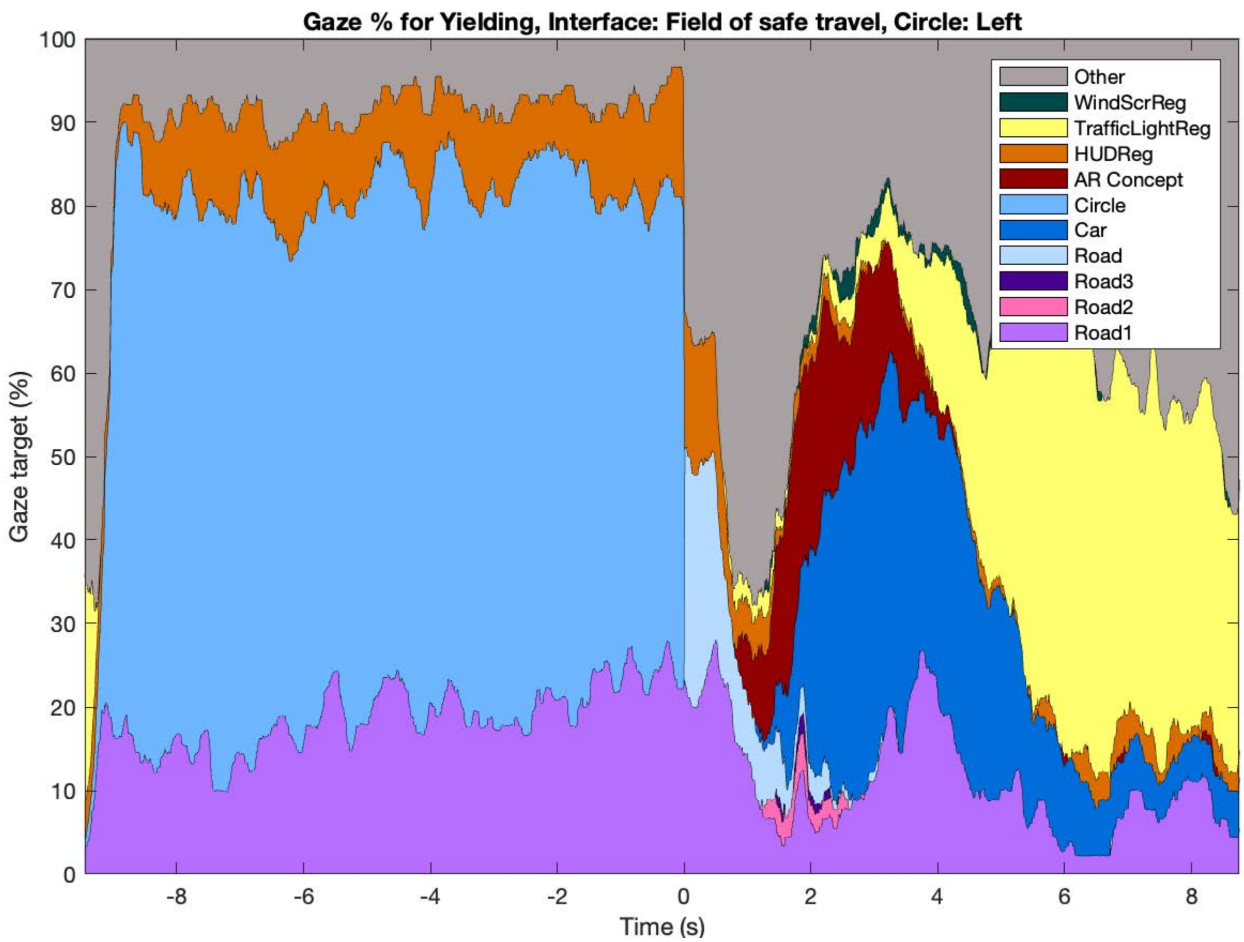
Figure 10. Bar plot of the moment in which the participants first glanced at the AR interface, as affected by the attention-attractor position, per AR interface condition. Vertical lines delineate the road-mapped interfaces, vehicle-mapped infaces, and HUDs. Out of a total of 3240 trials (9 AR interfaces × 12 trials × 30 participants), participants did not glance at the AR interface in 213 instances. This may potentially be attributed to inaccuracies in the eye-tracker, or to the genuine lack of attention directed toward the AR interface by the participants.

3.2.3 Distribution of Gaze across AOIs

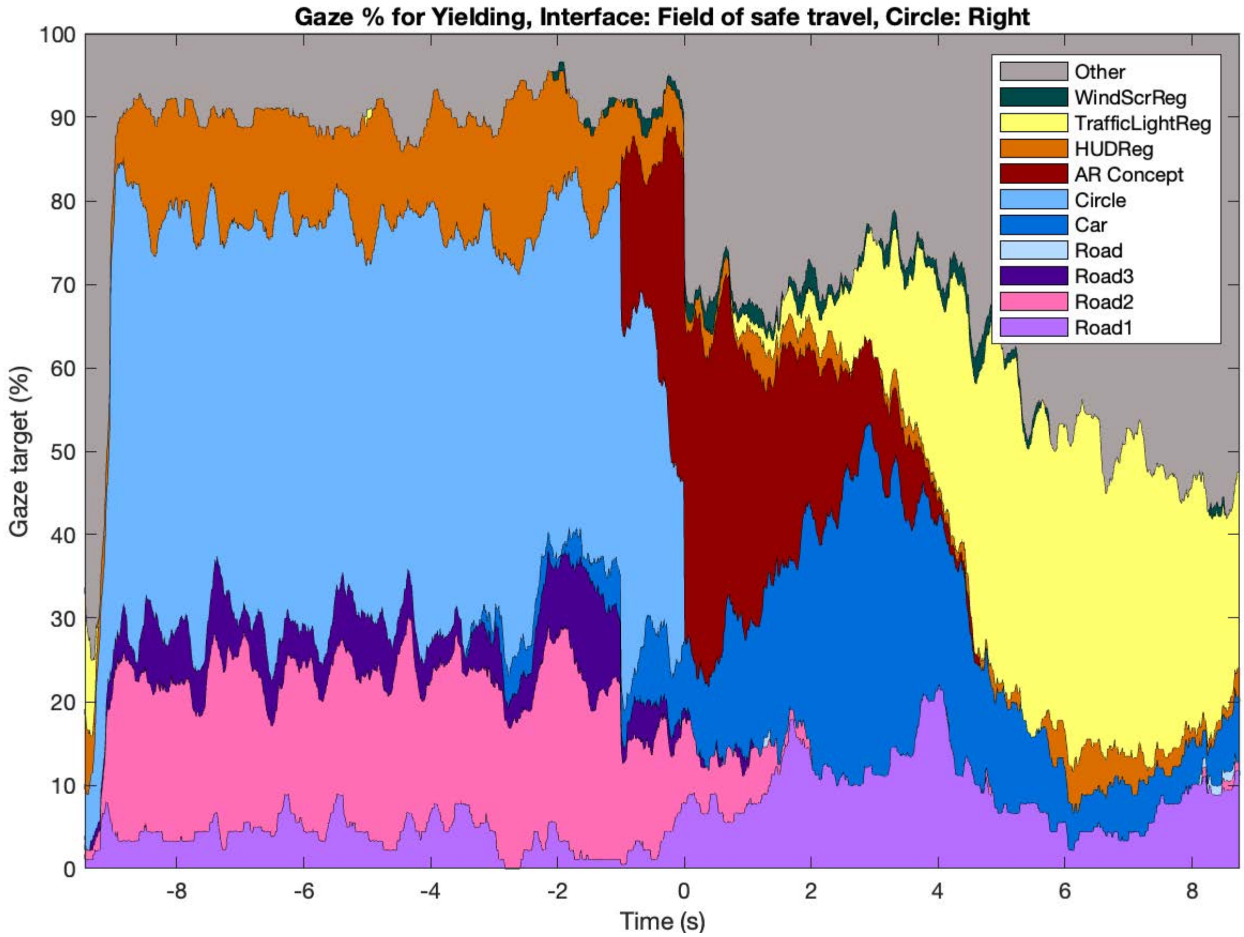
Figure 11 shows a sample of four gaze plots, pertaining to the yielding states of the *Field of safe travel* (a, b), and the *Pedestrian lights HUD* (c, d). In each case, the area plot depicts the gaze distribution across time for the different areas of interest. The two interfaces were chosen as they clearly demonstrate the effect of interface placement on the participants' attention distribution.

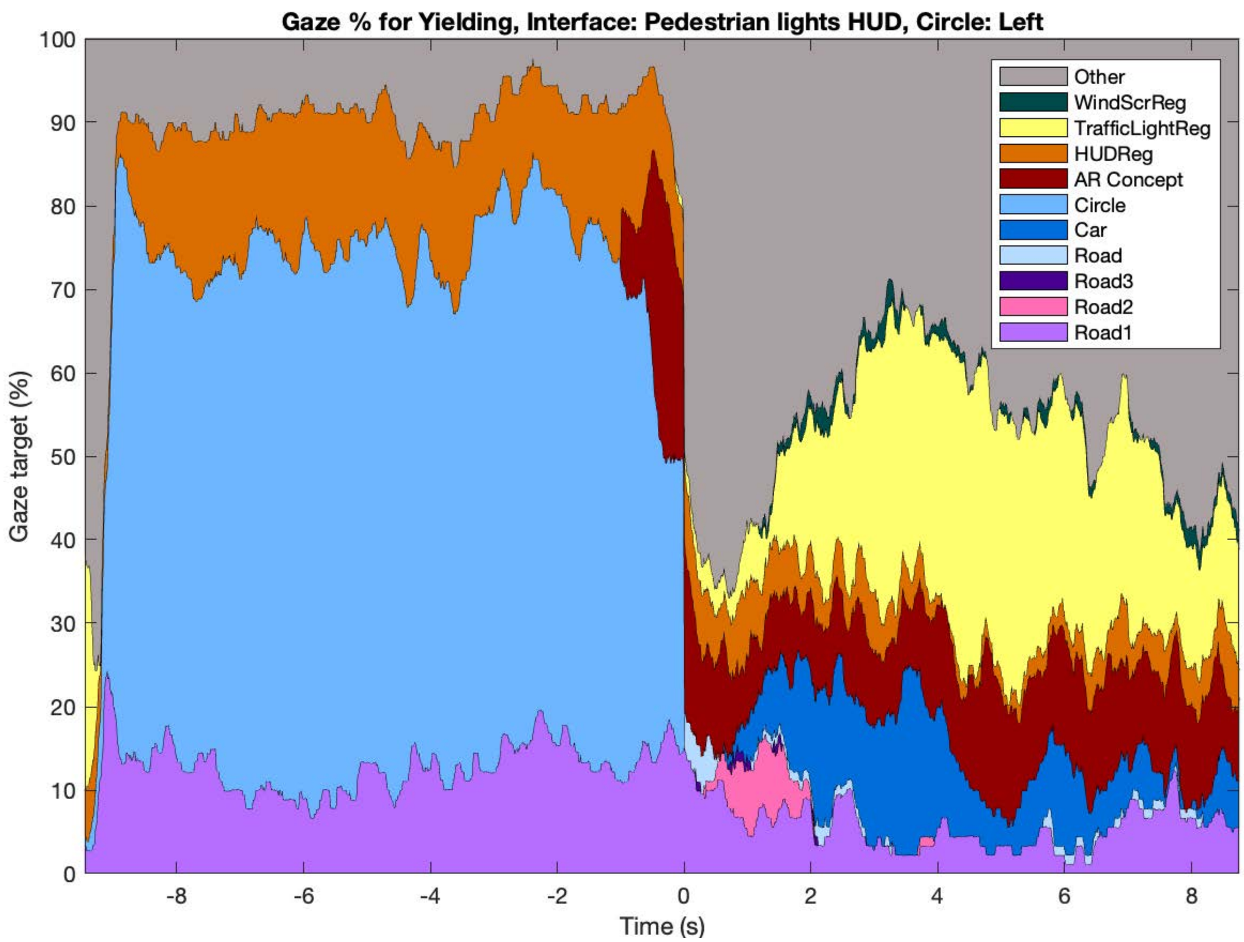
Figure 11 (a, c) shows the gaze distribution when the circle was presented on the left, while Figure 11 (b, d) corresponds to the circle presented on the right. When the circle was presented to the right, the participants were able to direct their gaze to the approaching vehicle ("Car") and the presented interface, i.e., *Field of safe travel* ("AR Concept") much earlier after time 0.0 s, compared to when the circle was presented to the left. In that case, there was a delay of over 1 second.

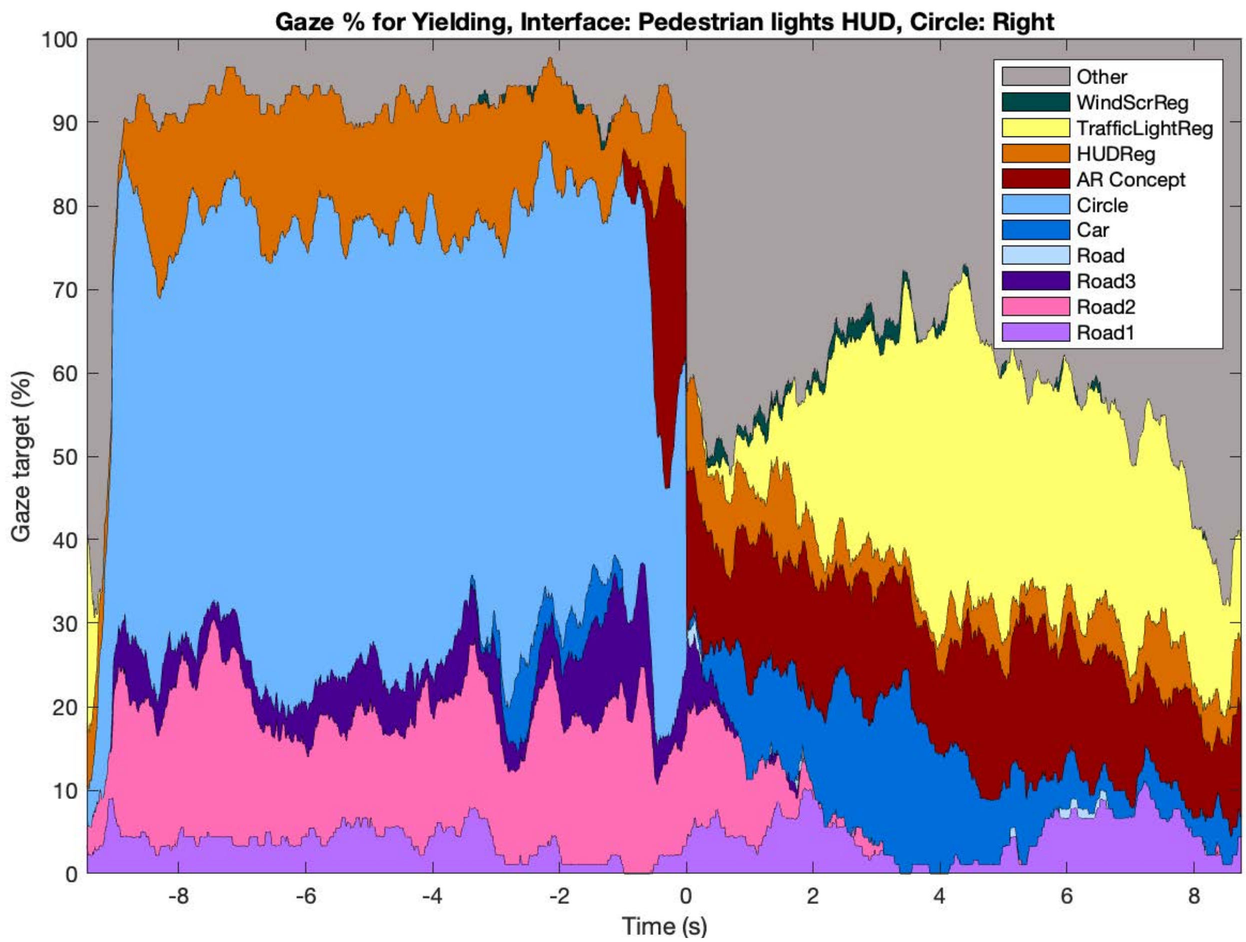
On the other hand, in the case of the *Pedestrian lights HUD*, the AR interface was gazed at before 0.0 s in both the left and right cases. These results confirm the advantage of the HUD, since it follows the user's gaze as a head-locked augmentation, whereas the *Field of safe travel* only has an advantage when a participant starts the trial while looking to the right.



(a)


Gaze % for Yielding, Interface: Field of safe travel, Circle: Right
Other
WindScrReg
TrafficLightReg
HUDReg
AR Concept
Circle
Car
Road
Road3
Road2
Road1
Gaze target (%)
Time (s)
100
90
80
70
60
50
40
30
20
10
0
-8
-6
-4
-2
0
2
4
6
8
(b)


Gaze % for Yielding, Interface: Pedestrian lights HUD, Circle: Left
Other
WindScrReg
TrafficLightReg
HUDReg
AR Concept
Circle
Car
Road
Road3
Road2
Road1
Gaze target (%)
Time (s)
100
90
80
70
60
50
40
30
20
10
0
-8
-6
-4
-2
0
2
4
6
8
(c)



(d)

Figure 11. Plots of the gaze percentage for each area of interest region in the environment. The figures above are for the yielding-state condition of the *Field of safe travel* interface, with the attention-attractor circle projected to the left (a), and right (b); and for the *Pedestrian lights HUD* with the attention-attractor circle projected to the left (c), and right (d). Time = 0 corresponds to the moment that the circle disappeared.

All the other gaze and head tracking plots for the other conditions can be found in the Supplementary Repository made available as a complement to this paper (see Appendix, section A.5).

3.3 Post-Block Interviews

Table 3 presents the outputs of the prompts from the Chat GPT-4 analysis, where the model was asked to output the strengths, and weaknesses of each interface based on 30 transcripts. The initial summaries can be found in the Supplementary Repository (link in Appendix, section A.5).

The overall summaries offered a fingerprint of what the participants said after each block. In general, interfaces mapped to the road (*Augmented zebra crossing, Fixed pedestrian lights, and Virtual fence*) were found to elicit confidence due to the participants' familiarity with common traffic designs. Moreover, the *Virtual fence*'s walls elicited a "sense of safety". On the other hand, the participants pointed out that such interfaces may cause users to focus too much on the design, rather than the surroundings, as the view behind the tunnel walls somewhat occludes the environment, despite being semi-translucent . Such familiar designs may also make the user overly reliant on the interface, or cause a false sense of security.

Interfaces mapped to the vehicle (*Planes on vehicle, Conspicuous looming planes, Field of safe travel, Phantom car*) were generally described as providing clear and distinguishable coloured cues. In fact, the use of looming in *Conspicuous looming planes* was described as effectively communicating urgency, and capturing the user's attention. However, these interfaces were also described as having the potential to not be immediately intuitive for the user, cause problems for colour-blind users in distinguishing the different states, confusion with animations (such as the scaling in *Conspicuous looming planes*), and ineffectiveness if not looking in the direction of the approaching vehicle. In general, participants felt that this group of interfaces had a large learning curve.

The HUD interfaces (*Nudge HUD, Pedestrian lights HUD*) were described as "providing a seamless experience" since the interface follows the user's gaze, which made the experience convenient, thus "boosting user confidence in crossing decisions". On the other hand, the HUDs were sometimes considered confusing, uncomfortable, and distracting, requiring a period of adjustment. Moreover, participants commented on the lack of customisation options for the HUD interfaces.

In the *Baseline* condition without any AR interface, participants stated that the interfaces helped provide "clear visual cues and signals that help users make informed decisions when crossing the road", which "can improve users' confidence and awareness while crossing". However, the participants also pointed out that users of the AR system might "become overly reliant on AR interfaces, which may decrease their ability to judge traffic situations independently", or the interface may not provide enough information or guidance in particular situations, which could ultimately cause confusion or hesitation.

Table 3.
Interview summaries by ChatGPT-4. In the summaries for the Baseline condition, an 'AR interface' is still alluded to. The reason is that the ChatGPT prompt only mentioned 'AR interfaces'.

Interface	Strengths	Weaknesses
Augmented zebra crossing	The AR interface is generally easy to understand, with clear and familiar signals such as green and red colours. It provides a sense of confidence and safety for users, effectively communicating when it is safe or unsafe to cross the road.	The interface may cause users to focus too much on the signals and not enough on their surroundings. It is less effective when users are not looking straight ahead, and the green signal is often less visible or intuitive than the red signal.
Planes on vehicle	The AR interface generally provides clear and easily distinguishable visual cues, such as colour coding and symbols, which aid users in making informed decisions about crossing	The AR interface may not be immediately intuitive for all users, requiring an adjustment period. Some participants found certain aspects of the interface, such as the red signal, less

	the road. The interface is often intuitive and can be learned over time, increasing user confidence and understanding of vehicle intentions.	effective or attention-grabbing. The interface may not account for all possible scenarios or adapt well to different contexts, potentially leading to confusion or misinterpretations.
Conspicuous looming planes	The AR interface effectively communicates urgency and safety through colour coding (red and green) and size changes, capturing users' attention and building trust over time. The interface is generally intuitive and easy to understand, with some users preferring it over other options.	The interface may not be immediately intuitive for first-time users, and its reliance on colour could be problematic for colorblind users or in situations with poor visibility. The green interface is often less effective and noticeable than the red one, and inconsistencies in design elements, such as scaling, can cause confusion and distraction for users.
Field of safe travel	The red signal in the AR interface is highly visible, easily understandable, and effective in conveying safety information. The colour-coded design is generally simple for users to comprehend, and the interface has the potential to improve pedestrian safety.	The green signal is less visible, less effective, and can cause confusion for users. The interface may require a learning curve for users to fully understand its purpose and meaning. The reliance on colour cues may not be effective for users with colour vision deficiencies, and the interface's effectiveness may vary in different lighting conditions or environments.
Fixed pedestrian lights	The AR interface is familiar, intuitive, and easy to understand, using recognizable traffic light symbols and colours. It provides clear signals and guidance for users to make informed decisions about crossing the road.	The interface may be distracting, difficult to notice, or require users to adjust their perspective, potentially impacting safety and decision-making. Users may become overly reliant on the interface, ignoring real-world cues, and the effectiveness of the interface may vary depending on lighting, weather conditions, and individual preferences.
Virtual fence	The AR interface effectively communicates safety through clear visual cues, such as red and green walls, and is generally intuitive and easy to understand. It provides a sense of safety and confidence for users when crossing the road and is adaptable to various situations.	The interface can be visually overwhelming and obstruct the view of vehicles, potentially causing confusion or a false sense of security. It may not cater to individual preferences or be suitable for colorblind users. Additionally, the reliance on the AR interface may lead to a lack of personal responsibility in making safe decisions while crossing the street.
Phantom car	The AR interface's colour coding (green for safe, red for danger) effectively communicates safety levels and aids decision-making. The red signs are particularly helpful and comfortable, providing insight for crossing the road. The interface can improve user alertness and engagement, potentially enhancing safety.	The interface can be initially confusing and may have a learning curve for new users. The green virtual vehicle is less intuitive and can be misleading, while the red car's display can be distracting and difficult to judge. The interface's effectiveness is dependent on the user's focus and direction, potentially causing confusion and reducing its overall utility.
Nudge HUD	The AR interface is generally intuitive, with clear colour coding (red and green) and text	The interface may cause initial discomfort or confusion, and users may still hesitate to trust it

	for guidance. It follows the user's eye movement, providing a seamless experience and boosting user confidence in crossing decisions.	fully, relying on real-world cues. There is room for improvement in terms of simplicity, distinct visual cues, and catering to different user preferences and needs.
Pedestrian lights HUD	The AR interface is generally clear, easy to understand, and familiar, using traffic light symbols and colours. It effectively communicates safety information and follows the user's gaze, making it convenient and efficient.	Some participants found the interface initially confusing, distracting, or disorienting. There is a lack of customization options and the interface may not be universally preferred. Trust in the interface varies, with some users preferring to rely on their own judgement or finding certain aspects (e.g., colours or symbols) less intuitive.
Baseline	The AR interface is generally user-friendly, providing clear visual cues and signals that help users make informed decisions when crossing the road. It simulates real-life traffic situations and can improve users' confidence and awareness while crossing.	Users may become overly reliant on the AR interface, potentially decreasing their ability to judge traffic situations independently. The interface may not provide enough information or guidance in certain situations, leading to confusion or hesitation. Additionally, the simplicity of the simulation may not accurately represent complex real-world scenarios, limiting the effectiveness of the training.

4. Discussion

Thirty participants completed 120 trials each in a CAVE pedestrian simulator experiment that assessed nine AR interfaces for pedestrian-vehicle interaction, and a baseline condition without any interface. The participants were asked to cross the virtual road if they felt it was safe to do so, and to rate how intuitive the interfaces were in communicating their message of intent. In addition to the standard head-tracker used in the CAVE, participants were equipped with an eye-tracker, which recorded their point of gaze during the experiment. Additionally, participants were asked about each interface in an audio-recorded interview following exposure. The data were analysed to assess the intuitiveness of each interface, the willingness to cross the road, the gaze and head direction of participants, and the further information provided during the interviews.

The first objective of the experiment was to evaluate the effect of nine AR interfaces on pedestrian crossing behaviour and perceived intuitiveness, with the expectation of replicating the intuitive ratings observed in a previous online questionnaire (Tabone et al., 2023). The second objective was to examine the effect of the position of the attention attractor on pedestrian crossing behaviour and perceived intuitiveness across different AR interfaces. An AR interface/ attention-position interaction was hypothesised, where head-locked, infrastructure-locked, and vehicle-locked AR interfaces would exhibit varying levels of intuitiveness based on the positioning of the attention-attractor circle.

4.1 Objective 1: Replication of Intuitiveness Ratings: CAVE study vs. Online Study

The present study was conducted in a highly immersive CAVE simulator, in which the participant (pedestrian) was surrounded by images that are displayed on the floor and walls (Cruz et al., 1992), while the scene orients according to the head position of the user. The CAVE allowed participants to see their own body (in contrast to online studies, or when using a HMD), allowing for a more natural interaction with the environment (Blissing & Bruzelius, 2018). In our experiment, none of the 30 participants suffered any discomfort related to motion sickness despite having gone through 120 trials each. Hence, when compared to HMDs or video-based study, a CAVE offers a more natural setting (Schneider et al., 2022).

Despite the high level of immersiveness offered by the CAVE, the AR interfaces achieved similar relative differences in mean intuitiveness scores as in the online questionnaire study. In fact, a correlation of $r = 0.91$ was observed between the two measures for yielding, and $r = 0.90$ for non-yielding, indicating that the intuitiveness scores highly replicated the online study. A correlation of 0.90 is strong, especially considering the limited sample size of only 30 participants, which implies imperfect statistical reliability. As in the online study, the interfaces which employed traditional traffic elements (*Augmented zebra crossing,* and *Fixed pedestrian lights)*, the HUDs (*Nudge HUD*, and *Pedestrian lights HUD*), and other interfaces mapped to the road (*Virtual fence)* were rated relatively highly, while interfaces mapped to the vehicle were deemed less intuitive (*Field of safe travel, Planes on vehicle, Conspicuous looming planes,* and *Phantom car*). The behavioural measure, i.e., crossing initiation time, also exhibited a substantial correlation with the online intuitiveness ratings ($r = -0.90$).

Unintended effects were once again observed for the *Field of safe travel*, and the *Phantom car,* which were designed to adhere to the principle of ecological interface design (Tabone, Lee, et al., 2021), and the principle of predictive aiding, respectively. However, similar to the online study, in the simulator study, the *Phantom car* may have failed to comply with the proximity compatibility principle, as participants found it difficult to separate the *Phantom car* from the 'real' car (as mentioned in Table 3). Moreover, both interfaces only displayed a coloured element without any icon or text, and may have therefore failed to comply with the principle of redundancy gain. Given that both interfaces lacked the incorporation of traditional design elements from traffic, the application of the top-down processing principle, which relies on the recognition of familiar symbols, may have posed another challenge for participants in terms of initially comprehending the designs intuitively. Conversely, the interfaces that achieved the highest intuitiveness scores demonstrated adherence to these aforementioned principles (Tabone et al., 2021).

The remarkably strong correlation between the outcomes of the online study and those of the CAVE simulator raises questions about the necessity of conducting experiments in a resource-intensive CAVE. Based on our results, we contend that if the objective is solely to gather subjective evaluations, such as average intuitiveness ratings, preference rankings, an online questionnaire could not only be sufficient but even preferable, considering the potential for larger sample sizes (refer tp Schneider et al., 2022).

Nonetheless, it is acknowledged that certain discrepancies emerged between the subjective impressions of the online study and the current CAVE-based study. For example, the HUD interfaces, which tracked participants' head motion, were experienced in real-time, whereas these motions were preprogrammed in the online study. This allowed participants to provide feedback on their functioning, which was occasionally described as jittery, annoying, or confusing. Moreover, the *Virtual fence* might create a false sense of security and potentially obstruct the view of oncoming vehicles, which are features that could not be observed in the prior online study, which used video clips. Similarly, in the preference rankings, the *Virtual fence* received a notably high number of first-place rankings, which could be attributed to its visually appealing animation of opening doors and its comprehensive coverage, potentially appealing to participants due to a "coolness" factor.

4.2 Objective 2: AR interface/Attention-Position Interaction

A main reason the study was conducted in the HIKER was due to its high field-of-view compared to an online experiment, allowing for the guidance of the participant's initial attention allocation. An attention-attractor circle was utilised to guide the participant's attention towards a specific region of the road environment before the arrival of the vehicle, as in real-world situations, pedestrians may initially neglect to observe the approaching vehicle due to finite attentional resources (Ralph & Girardeau, 2020; Wickens et al., 2004). Hence, the effect of the position of the attention-attractor circle on pedestrian crossing behaviour and perceived intuitiveness across different AR interfaces could be explored.

The position of the attention-attractor circle did in fact have an effect on the crossing initiation time of the participants. As detailed in Figure 9, there was a significant delay in crossing initiation when participants were presented with an interface mapped to the vehicle while the circle was presented to the left.

Interfaces mapped to the road, in particular the *Fixed pedestrian lights*, registered a faster crossing initiation time when the attention-attractor circle was presented in the centre position, while the HUDs had similar initiation times for each of the three attention-attractor positions. For the HUDs, this could be explained by the fact that it remained in the central field-of-view of the participant, regardless of the participant's head orientation. On the other hand, the outcomes for the road-mapped interfaces were supported by the post-trial interviews, where the *Augmented zebra crossing* was found to be "less effective when users are not looking straight ahead". A surprising outcome from the interviews was that the HUDs received criticism for 'lack of customisation' and were also found to be 'distracting and disorienting'. The unintended effect was revealed because, in the CAVE, it was possible to have the interfaces follow the participant's gaze in contrast to video-based studies. However, the positive comments on the HUDs did confirm the advantage they offer in following the users' gaze.

The *Virtual fence* was also unaffected by the attention-attractor position. Two possible explanations can account for this observed phenomenon. Firstly, the *Virtual fence* was large and visually conspicuous, thereby increasing the likelihood that participants covertly perceive its presence and colour through peripheral vision, even if they do not directly glance at it. The size of the *Virtual fence* was also mentioned by participants in the interviews. Secondly, the substantial size of the virtual fence made it more probable that participants could rapidly direct their foveal vision toward it. This assertion is supported by the findings illustrated in Figure 10 (time of first glance) and Figure 9 (crossing initiation times), which demonstrated that the *Virtual* *fence* performed well in these regards. In contrast, other AR interfaces were smaller, such as the *Fixed pedestrian lights*, or moving within the scene, such as the vehicle-locked interfaces, rendering them more challenging to detect with peripheral vision (i.e., before turning the eyes) and foveal vision (i.e., by moving the eyes).

The inferior performance of the interfaces mapped to the vehicle may have implications for eHMI design, taking into account that proposed eHMIs have mostly been placed on the vehicle itself. On the other hand, the performance of the HUDs, being less affected by user distraction, makes them ideal candidates for use in such traffic interactions, because in everyday situations, the pedestrian cannot be expected to look at an approaching vehicle right away. More so, they are also the easiest to implement in terms of sensor requirements, since the HUDs would need minimal context from the environment (i.e., only AV intent), unlike some other interfaces, such as the *Planes on Vehicle*, which in reality would need to make use of computer vision methods to detect the vehicle (Tabone et al., 2021b). At the same time, it may be important to consider the potential risk of overreliance. A more detailed exploration of the eye-tracking data revealed that in the Baseline condition, participants neglected to glance at the AV in merely 9 out of 360 trials. In contrast, for the *Augmented Zebra Crossing*, *Fixed Pedestrian Lights*, *Nudge HUD*, and *Pedestrian Lights HUD*, the respective numbers were 76, 71, 70, and 72 out of 360. This observation indicates that the AR interfaces could potentially foster crossing behaviour without proper looking (for a similar concern in eHMI research, see Kaleefathullah et al., 2022).

An intriguing observation emerged regarding the crossing initiation times for the Baseline condition, which appeared to be only slightly affected by the position of the attention attractor. A plausible explanation for this phenomenon is that, although the attention attractor positioned on the left and centre caused a delayed response in glancing at the approaching AV (just as it did for the nine AR interface conditions), this temporal delay is not detrimental. In the Baseline condition, no valuable information would be missed, as it is challenging to discern whether a car has initiated braking from implicit communication (i.e., vehicle speed) alone. The AR interfaces, on the other hand, were activated 0.8 seconds *before* the AV started to slow down. Therefore, a pedestrian would miss valuable explicit communication if they do not glance promptly at the AR interface (and see De Winter & Dodou, 2022, for a discussion on eHMIs that provide anticipatory information).

4.3 Limitations and Future Work

Despite the enhanced ecological validity when compared to the online questionnaire study in Tabone et al. (2023a), there are still some limitations to the study presented in this paper. Firstly, the CAVE and the equipment itself were sources of distraction as there were points during the interviews where the participant was commenting about the VR environment rather than the interfaces themselves.

Other limitations arose from equipment issues. The eye-tracker had a limited battery life, and there were instances when the battery needed to be replaced between blocks. Moreover, on average, 4 or 5 participants were accommodated in a single experiment day, resulting in a cumulative running time of over 9 to 10 hours. This extended duration occasionally led to equipment overheating. Also, the *Field of Safe Travel* and *Phantom Car* interfaces experienced some rubberbanding issues, which manifested as jittery and inconsistent depth perceptions. This phenomenon may have impacted the reported intuitiveness ratings to a certain degree. In fact, some post-block interviews (such as P20 and P24; see Supplementary Material) for the *Phantom Car* reported instances of abnormal behaviour, including “one of the green cars reversed” and “on a couple of cases with the green, the car went past the hologram.” These issues were due to occasional synchronisation problems across screens and computers, as well as limitations of the Unity networking system. Regardless of these issues, a high correlation between the simulator and online questionnaire results was observed.

Additionally, the limited screen area of the CAVE meant that some of the interfaces had to be scaled down to fit. In fact, the size of both the HUD interfaces were reduced from their original dimension during piloting as they became too large to fit on the CAVE screens as the participant approached the wall across (end of the virtual road). In fact, in the post-experiment subjective ranking results there were some comments about the HUD interfaces being ‘too small’. Such an unintended effect was therefore caused by technical limitations of a CAVE simulator.

The AR interfaces were still projected on a virtual environment, which ultimately rendered them as part of the VR simulation. Furthermore, despite the CAVE simulation being immersive, it still did not offer real risk, and participants were asked to cross repeatedly without any real purpose other than to complete the experiment. Hence, to better understand the trust and behaviour of participants with these AR interfaces, the ecological validity must once again be increased and the AR interfaces be tested overlaid over the real world layer.

5. Conclusion

A pedestrian simulator experiment was conducted to investigate the interaction between pedestrians and automated vehicles using nine different augmented reality interfaces presented within a virtual reality CAVE. The aim of the experiment was twofold: (1) to investigate whether results from an online questionnaire video-based study could be replicated in a CAVE simulator, and (2) to investigate whether the effectiveness of different AR designs is contingent upon the pedestrian’s attention allocation. The statistical and qualitative findings indicate that the mean intuitiveness ratings correlated substantially with results from a prior online study ($r \approx 0.90$). Interfaces using traditional traffic design elements, and head-locked HUDs were rated as more intuitive, in contrast to vehicle-mapped interfaces.

The position of the attraction-attractor circle affected crossing initiation time, with participants initiating crossing earlier for road-mapped and HUD interfaces in general, in contrast to interfaces that were mapped to the vehicle. For the *Virtual fence* and the HUDs, crossing initiation times were relatively unaffected by the attention-attractor position. These findings are likely due to the size of the *Virtual fence* and the omnipresence of the HUD interfaces in the participant’s FOV.

The poor performance of vehicle-mapped AR interfaces may have implications for eHMIs, which are typically mounted on the vehicle. Lastly, the experiment outcomes also highlight the fact that a CAVE simulator might not be necessary if the goal is to investigate intuitiveness ratings, given the high correlation with the online study. However, the immersive nature of the CAVE proved essential for studying the distributed attention of the participants.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860410.

References

Alam, F., Mehmood, R., Katib, I., Albogami, N. N., & Albeshri, A. (2017). Data fusion and IoT for smart ubiquitous environments: A survey. IEEE Access, 5, 9533–9554. https://doi.org/10.1109/ACCESS.2017.2697839

Arena, F., Collotta, M., Pau, G., & Termine, F. (2022). An overview of augmented reality. *Computers*, *11*(2), 28. https://doi.org/10.3390/computers11020028

Banks, V. A., Stanton, N. A., Burnett, G., & Hermawati, S. (2018). Distributed Cognition on the road: Using EAST to explore future road transportation systems. Applied Ergonomics, 68, 258–266. https://doi.org/10.1016/j.apergo.2017.11.013

Bazilinskyy, P., Dodou, D., & De Winter, J. C. F. (2019). Survey on eHMI concepts: The effect of text, color, and perspective. Transportation Research Part F: Traffic Psychology and Behaviour, 67, 175–194. https://doi.org/10.1016/j.trf.2019.10.013

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

Bhorkar, G. (2017). A survey of augmented reality navigation. arXiv. https://arxiv.org/abs/1708.05006

Blissing, B., & Bruzelius, F. (2018). Exploring the suitability of virtual reality for driving simulation. In Driving Simulation Conference 2018 (pp. 163–166).

Bos, J. E., MacKinnon, S. N., & Patterson, A. (2005). Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. Aviation, space, and environmental medicine, 76, 1111–1118.

Calvi, A., D'Amico, F., Ferrante, C., & Ciampoli, L. B. (2020). Effectiveness of augmented reality warnings on driving behaviour whilst approaching pedestrian crossings: A driving simulator study. Accident Analysis & Prevention, 147, 105760. https://doi.org/10.1016/j.aap.2020.105760

Cao, D., Wang, X., Li, L., Lv, C., Na, X., Xing, Y., Li, … & Wang, F. Y. (2022). Future directions of intelligent vehicles: Potentials, possibilities, and perspectives. IEEE Transactions on Intelligent Vehicles, 7, 7–10. https://doi.org/10.1109/TIV.2022.3157049

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia tools and applications*, *51*, 341–377. https://doi.org/10.1007/s11042-010-0660-6

Colley, M., Eder, B., Rixen, J. O., & Rukzio, E. (2021). Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445351

Cordeil, M., Cunningham, A., Dwyer, T., Thomas, B. H., & Marriott, K. (2017). ImAxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In Proceedings of the 30th annual ACM symposium on user interface software and technology, 71–83. https://doi.org/10.1145/3126594.3126613

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. Communications of the ACM, 35, 64–73. https://doi.org/10.1145/129888.129892

Currano, R., Park, S. Y., Moore, D. J., Lyons, K., & Sirkin, D. (2021). Little road driving hud: Heads-up display complexity influences drivers' perceptions of automated vehicles. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan. https://doi.org/10.1145/3411764.3445575

Dancu, A., Vechev, V., Ünlüer, A. A., Nilson, S., Nygren, O., Eliasson, S., Barjonet, J.-E., Marshall, J., & Fjeld, M. (2015). Gesture bike: Examining projection surfaces and turn signal systems for urban cycling. Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, Madeira, Portugal, 151–159. https://doi.org/10.1145/2817721.2817748

De Clercq, K., Dietrich, A., Núñez Velasco, J. P., De Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Human Factors*, *61*, 1353–1370. https://doi.org/10.1177/00187208198363

De Winter, J., & Dodou, D. (2022). External human-machine interfaces: Gimmick or necessity. Transportation Research Interdisciplinary Perspectives, 15, 100643. https://doi.org/10.1016/j.trip.2022.100643

Dey, D., Habibovic, A., Löcken, A., Wintersberger, P., Pfleging, B., Riener, A., Martens, M., & Terken, J. (2020a). Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. Transportation Research Interdisciplinary Perspectives, 7, 100174. https://doi.org/10.1016/j.trip.2020.100174

Dong, W., Wu, Y., Qin, T., Bian, X., Zhao, Y., He, Y., Xu, Y., & Yu, C. (2021). What is the difference between augmented reality and 2D navigation electronic maps in pedestrian wayfinding? Cartography and Geographic Information Science, 48, 225–240. https://doi.org/10.1080/15230406.2021.1871646

Ginters, E. (2019). Augmented reality use for cycling quality improvement. Procedia Computer Science, 149, 167–176. https://doi.org/10.1016/j.procs.2019.01.120

Hesenius, M., Börsting, I., Meyer, O., & Gruhn, V. (2018). Don't panic! Guiding pedestrians in autonomous traffic with augmented reality. Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, Barcelona, Spain, 261–268. https://doi.org/10.1145/3236112.3236148

Ishihara, S. (1917). Tests for colour-blindness. Tokyo: Kanehara Shuppan Co., Ltd.

Jerald, J. (2016). The VR book: Human-centered design for virtual reality. Morgan & Claypool.

Kalantari, A. H., Yang, Y., Merat, N., & Markkula, G. (2023). Modelling vehicle-pedestrian interactions at unsignalised locations: Road users may not play the Nash equilibrium. Manuscript submitted for publication.

Kaleefathullah, A. A., Merat, N., Lee, Y. M., Eisma, Y. B., Madigan, R., Garcia, J., & De Winter, J. C. F. (2022). External human–machine interfaces can be misleading: An examination of trust development and misuse in a CAVE-based pedestrian simulation environment. Human Factors, 64, 1070–1085. https://doi.org/10.1177/0018720820970751

Kadar, E. E., & Shaw, R. E. (2000). Toward an ecological field theory of perceptual control of locomotion. *Ecological Psychology*, *12*, 141–180. https://doi.org/10.1207/S15326969ECO1202_02

Kim, H., Gabbard, J. L., Anon, A. M., & Misu, T. (2018). Driver behavior and performance with augmented reality pedestrian collision warning: An outdoor user study. IEEE Transactions on Visualization and Computer Graphics, 24, 1515–1524. https://doi.org/10.1109/TVCG.2018.2793680

Lanzer, M., Koniakowsky, I., Colley, M., & Baumann, M. (2023). Interaction Effects of Pedestrian Behavior, Smartphone Distraction and External Communication of Automated Vehicles on Crossing and Gaze Behavior. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3544548.3581303

Lee, J. D. (2008). Fifty years of driving safety research. Human Factors, 50, 521–528. https://doi.org/10.1518/001872008X2883

Li, T., Guo, F., Krishnan, R., Sivakumar, A., & Polak, J. (2020). Right-of-way reallocation for mixed flow of autonomous vehicles and human driven vehicles. Transportation research part C: emerging technologies, 115, 102630. https://doi.org/10.1016/j.trc.2020.102630

Li, Y., Kamalasanan, V., Batista, M., & Sester, M. (2022). Improving pedestrian priority via grouping and virtual lanes. arXiv. https://arxiv.org/abs/2205.08783

Mallaro, S., Rahimian, P., O'Neal, E. E., Plumert, J. M., & Kearney, J. K. (2017). A comparison of head-mounted displays vs. large-screen displays for an interactive pedestrian simulator. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. https://doi.org/10.1145/3139131.3139171

Matviienko, A., Müller, F., Schön, D., Seesemann, P., Günther, S., & Mühlhäuser, M. (2022). BikeAR: Understanding cyclists’ crossing decision-making at uncontrolled intersections using Augmented Reality. Proceedings of the CHI Conference on Human Factors in Computing Systems, New Orleans, LA. https://doi.org/10.1145/3491102.3517560

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Mukhopadhyay, A., Sharma, V. K., Tatyarao, P. G., Shah, A. K., Rao, A. M., Subin, P. R., & Biswas, P. (2023). A comparison study between XR interfaces for driver assistance in take over request. Transportation Engineering, 11, 100159. https://doi.org/10.1016/j.treng.2022.100159

Otter.ai (2023). Otter.ai - Voice meeting notes & real-time transcription. https://otter.ai

Pichen, J., Yan, F., & Baumann, M. (2020). Towards a cooperative driver-vehicle interface: Enhancing drivers’ perception of cyclists through Augmented Reality. Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, 1827–1832. https://doi.org/10.1109/IV47402.2020.9304621

Pratticò, F. G., Lamberti, F., Cannavò, A., Morra, L., & Montuschi, P. (2021). Comparing state-of-the-art and emerging Augmented Reality interfaces for autonomous vehicle-to-pedestrian communication. IEEE Transactions on Vehicular Technology, 70, 1157–1168. https://doi.org/10.1109/TVT.2021.3054312

Qualtrics. (2022). Qualtrics XM // The leading experience management software. https://www.qualtrics.com

Ralph, K., & Girardeau, I. (2020). Distracted by "distracted pedestrians"?. Transportation research interdisciplinary perspectives, 5, 100118. https://doi.org/10.1016/j.trip.2020.100118

Rasouli, A., & Tsotsos, J. K. (2020). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Transactions on Intelligent Transportation Systems, 21, 900–918. https://doi.org/10.1109/TITS.2019.2901817

Rouchitsas, A., & Alm, H. (2019). External human–machine interfaces for autonomous vehicle-to-pedestrian communication: A review of empirical work. Frontiers in Psychology, 10, 2757. https://doi.org/10.3389/fpsyg.2019.02757

Saffer, D. (2010). *Designing for interaction: Creating innovative applications and devices*. Berkeley, CA: New Riders.

Sewalkar, P., & Seitz, J. (2019). Vehicle-to-pedestrian communication for vulnerable road users: Survey, design considerations, and challenges. Sensors, 19, 358. https://doi.org/10.3390/s19020358

Schneider, S., Maruhn, P., Dang, N. T., Pala, P., Cavallo, V., & Bengler, K. (2022). Pedestrian crossing decisions in virtual environments: behavioral validity in CAVEs and head-mounted displays. Human factors, 64, 1210–1226. https://doi.org/10.1177/0018720820987446

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, Leeds, UK, 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

Tabone, W., & De Winter J. C. F. (2023). Using ChatGPT for Human-Computer Interaction Research: A Primer. Manuscript submitted for publication.

Tafidis, P., Farah, H., Brijs, T., & Pirdavani, A. (2022). Safety implications of higher levels of automated vehicles: a scoping review. Transport reviews, 42, 245–267. https://doi.org/10.1080/01441647.2021.1971794

Toh, C. K., Sanguesa, J. A., Cano, J. C., & Martinez, F. J. (2020). Advances in smart roads for future smart cities. Proceedings of the Royal Society A, 476, 20190439. https://doi.org/10.1098/rspa.2019.0439

Tong, Y., & Jia, B. (2019). An Augmented-Reality-based warning interface for pedestrians: User interface design and evaluation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63, 1834–1838. https://doi.org/10.1177/1071181319631413

Tran, T. T. M., Parker, C., Wang, Y., & Tomitsch, M. (2022). Designing wearable augmented reality concepts to support scalability in autonomous vehicle–pedestrian interaction. Frontiers in Computer Science, 4, 866516. https://doi.org/10.3389/fcomp.2022.866516

Tran, T. T. M., Parker, C., Hoggenmüller, M., Hespanhol, L., & Tomitsch, M. (2023). Simulating wearable urban augmented reality experiences in VR: Lessons learnt from designing two future urban interfaces. Multimodal Technologies and Interaction, 7, 21. https://doi.org/10.3390/mti7020021

Unity. (2022). Unity real-time development platform | 3D, 2D VR & AR engine. https://www.unity.com

University of Leeds. (2022). Highly Immersive Kinematic Experimental Research (HIKER) pedestrian lab. https://uolds.leeds.ac.uk/facility/hiker-lab

Von Sawitzky, T., Wintersberger, P., Löcken, A., Frison, A. K., & Riener, A. (2020). Augmentation concepts with HUDs for cyclists to improve road safety in shared spaces. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI. https://doi.org/10.1145/3334480.3383022

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). Crossing the Street Across the Globe: A Study on the Effects of eHMI on Pedestrians in the US, Germany and China. In HCI in Mobility, Transport, and Automotive Systems: First International Conference, MobiTAS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, 21, 515–530. Springer International Publishing. https://doi.org/10.1007/978-3-030-22666-4_37

Wickens, C. D., Gordon, S. E., & Liu, Y. (2004). *An introduction to human factors engineering*. Upper Saddle River, NJ: Pearson Prentice Hall.

World Health Organisation. (2020). The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Appendix

A1. Demographics questionnaire

A preliminary question (Q0) was added to record the participant number, which was manually filled out by the experimenter beforehand. The questionnaire began by asking about the participant's identified gender (Q1), age (Q2), nationality (Q3), the length of time they had lived in the UK (Q4), and their familiarity with traffic systems (i.e., left-hand, right-hand, or both) (Q5), along with their occupation. The subsequent question (Q6) assessed the participant’s affinity for technological systems, using the Affinity for Technology Interaction (ATI) scale (Franke et al., 2019). This was followed by questions regarding previous use of VR headsets (Q7), AR apps (Q8), and participation in CAVE-based simulator experiments (Q9). Participants were then asked about their primary mode of transportation (Q10) and whether they possessed a driving licence (Q11). If they answered “yes” to the latter, two additional questions about the year they obtained their licence (Q12) and the number of miles driven in the past 12 months (Q13) were presented. The final question in the demographic questionnaire included a colour blindness test (Q14) (Ishihara, 1917; as used in Bazilinskyy et al., 2020 and Tabone et al., 2023).

A2. General characteristics of the participants

General characteristics of the 30 participants were as follows:

- 26.7% (*n* = 8) indicated that they had lived in the UK for 1–5 years, 3.3% (*n* = 1) lived in the UK between 5 and 10 years, 30% (*n* = 9) for less than a year, and 40% (*n* = 12) for more than 10 years.
- 33.3% (*n* = 10) were familiar with both left-hand traffic (LHT) and right-hand traffic (RHT), 43.3% (*n* = 13) indicated that they were familiar with only LHT, and 23.3% (*n* = 7) with only RHT.
- The mean total Affinity for Technology Interaction (ATI) scale score was 4.25.
- 26.7% (*n* = 8) had never used VR, while 73.3% (*n* = 22) indicated that they had used VR.
- 43.3% (*n* = 13) had never used AR, while 56.7% (*n* = 17) did.
- 56.7% (*n* = 17) had never been in a CAVE, while 43.3% (*n* = 13) indicated that they did.
- 50.0% (*n* = 15) indicated that their daily walking ranged between 15 and 30 minutes, 10.0% (*n* = 3) stated 30–45 minutes, 23.3% (*n* = 7) indicated 45–60 minutes, and 16.7% (*n* = 5) stated 60 minutes and above.
- For primary transportation, 13.3% (*n* = 4) use cycling, 23.3% (*n* = 7) use private vehicles, 16.7% (*n* = 5) use public transportation, and 46.7% (*n* = 14) mostly walk.
- 86.7% (*n* = 26) possess a driving licence, while 13.3% (*n* = 4) do not.
- From those that had a driving licence, 77% (*n* = 20) obtained it in the last 10 years, while 23% (*n* = 6) before that, with 1993 being the earliest year of attainment, and 2020 the most recent.
- 34.6% (*n* = 9) of participants with a licence, drive everyday, 3.8% (*n* = 1) drive just in the weekdays, 7.7% (*n* = 2) drive just in the weekends, 26.9% (*n* = 7) never drive, 7.7% (*n* = 2) drive once per week, 11.5% (*n* = 3) drived only once, and 7.7% (*n* = 2) preferred not to respond.

- From the licensed sample, 11.5% (*n* = 3) had driven 0 miles in the preceding 12 months, 38.5% (*n* = 10) had driven 1–5,000 miles, 11.5% (*n* = 3) had driven 10,001–15,000 miles, 7.7% (*n* = 2) had driven 15,001–20,000 miles, 23.1% (*n* = 6) had driven 5,001–10,000 miles, and 7.7% (*n* = 2) preferred not to respond.
- 3.3% (*n* = 1) of the respondent pool was considered colourblind as they had submitted three or more incorrect responses (Bazilinskyy et al., 2020; and Tabone et al., 2023) for the six-item Ishihara colour blindness test.
- During the experiment, only one participant from 30 stated that they were uncomfortable at one point in the experiment. The MISC scale was presented to the participant, but they had voted '1', which corresponds to 'some discomfort, but no specific symptoms' (Bos et al., 2005). Therefore, the experiment continued, as the score was not above 3.

A3. Scores of Dependent Measures

Table A3.1: Means of dependent measures or the nine AR interfaces and Baseline Condition

	1. Hiker: Intuitiveness (non-yielding vehicle)	2. Hiker: Intuitiveness (yielding vehicle)	3. Hiker: ChatGPT sentiment score	4. Hiker: Mean preference rank	5. Hiker: Crossing initiation time	6. Online: Intuitiveness (non-yielding vehicle)	7. Online: Intuitiveness (yielding vehicle)
Augmented zebra	5.98	6.16	81.82	4.34	2.71	5.67	5.53
Planes on vehicle	5.44	5.98	79.76	4.41	3.50	4.61	4.85
Conspicuous looming planes	5.56	5.71	76.37	5.62	3.55	4.64	4.65
Field of safe travel	6.10	5.56	76.76	5.07	3.67	4.95	4.84
Fixed pedestrian lights	6.32	6.23	80.54	5.03	3.00	5.62	5.57
Virtual fence	5.79	6.22	80.60	4.34	2.51	5.00	5.48
Phantom car	4.50	4.92	69.20	7.14	3.60	4.07	4.29
Nudge HUD	6.22	6.15	81.76	4.41	2.40	5.68	5.69
Pedestrian lights HUD	6.14	6.31	82.71	4.62	2.58	5.59	5.50
Baseline	5.48	4.80	77.36	—	4.69	—	—

Table A3.2: Post-experiment questionnaire: Interface rankings

Score	**Augmented zebra crossing**	**Planes on vehicle**	**Conspicuous looming planes**	**Field of safe travel**	**Fixed pedestrian lights**	**Virtual fence**	**Phantom car**	**Nudge HUD**	**Pedestrian traffic lights HUD**
1	5	3	3	1	4	**10**	0	3	0
2	5	3	3	4	0	1	1	**6**	**6**
3	2	**8**	1	2	2	2	1	5	6
4	6	2	2	6	**8**	1	1	1	2
5	1	**5**	3	**5**	1	**5**	1	4	4
6	2	2	4	1	5	1	5	3	**6**
7	2	1	5	4	4	2	**6**	3	2
8	4	2	3	**5**	4	4	**5**	1	1
9	2	3	**5**	1	1	3	**9**	3	2

A4. Supplementary Material

In the spirit of open science, and to assist the replicability of this study, a repository has been made available at (link following publication). It contains supplementary data to this paper, such as the processed comma-separated-value files, interview transcripts, ChatGPT outputs, gaze and head-tracking plots, and the code used to generate the results presented in this paper.