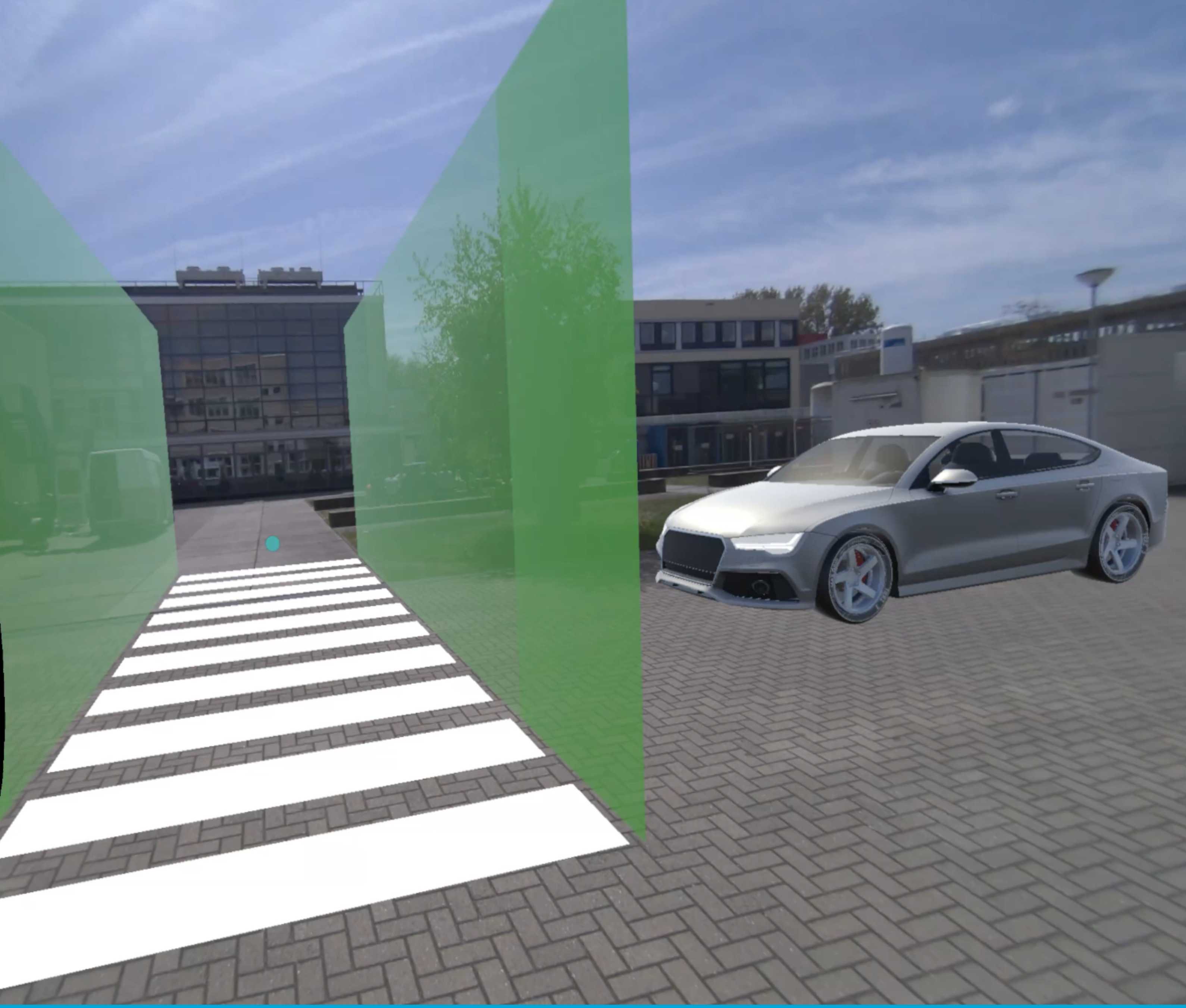

Evaluating Augmented Reality Interfaces for Pedestrians by conducting AR-based experiments in the Real World

Thomas Aleva

TUDelft

Evaluating Augmented Reality Interfaces for Pedestrians in the Real World by conducting AR-based experiments

Master's thesis by

Thomas Aleva

In fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at

Delft University of Technology

July 17th, 2023

Student number: 4588916

Supervised by:

Prof.dr.ir. J.C.F. (Joost) de Winter	TU Delft, supervisor
Dr. D. (Dimitra) Dodou	TU Delft, supervisor
W. (Wilbert) Tabone, MSc	TU Delft, daily supervisor

Preface

I am excited to present my Master's thesis, which explores autonomous-vehicle-to-pedestrian (AV2P) communication through augmented reality (AR) interfaces and its impact on pedestrian road-crossing behaviour. The motivation behind this study originated from the observation that prior research into AV2P communication interfaces have primarily relied on virtual reality (VR) settings, resulting in limited ecological validity and impacting participants' behaviour. Therefore, an AR experiment has been conducted in the real world to increase ecological validity and investigate pedestrian crossing behaviour.

The future potential of AV2P communications systems being commonly implemented, through consumer electronics such as Apple's Vision Pro, and improving traffic safety, provided me with personal motivation to conduct this study. Furthermore, as a regular driver, I experience how inefficient traffic interactions can be, which emphasises the importance of researching and developing systems that improve traffic safety and efficiency.

The experiments conducted for this study included a Varjo XR-3 head-mounted display, through which a virtual vehicle was projected, and four different interfaces (in addition to a no-interface baseline) were utilised to communicate the vehicle's intentions to the participants. Their *Willingness to cross* and *Decision certainty* were recorded through a button press, and subjective data was gathered to compare participants' perception of the AR experiment to a VR experiment.

Findings indicate a positive effect of AV2P interfaces on participants' Willingness to cross and Decision certainty. Furthermore, the comparison between subjective data from the AR experiments and an earlier VR study underline the importance of increased ecological validity.

I extend my gratitude to Joost de Winter, Dimitra Doudou, and Wilbert Tabone who supported and guided me throughout this journey providing invaluable advice, encouragement, feedback and support. I hope that my thesis provides useful knowledge that can be used to further research AV2P communication systems, and specifically contribute to Wilbert's PhD study. I am also grateful to the participants who voluntarily participated in the experiments, enabling the collection of valuable data. Their involvement has been paramount to the findings presented in this thesis.

I hope this thesis provides you with novel insights and an enjoyable read.

Thomas Aleva
Delft, July 16th 2023

Abstract

Objective: The aim of this study was to investigate the effect of autonomous-vehicle-to-pedestrian (AV2P) communication through augmented reality (AR) interfaces on the road crossing behaviour of pedestrians, and research whether subjective results from a previous Cave Automatic Virtual Environment (CAVE) study replicated in a real world AR experiment.

Background: Previous studies investigating the effects of AV2P communication have mostly been conducted through virtual reality (VR) providing researchers with safe experimentation methods and high experimental control, but also resulting in a common limitation: the lack of ecological validity and realism, thereby affecting participants' behaviour and causing distractions. This study therefore introduces AR experiments that have been conducted in a real world environment to increase ecological validity.

Methods: An AR experiment was conducted in which 28 participants were situated in the real world with the objective to cross the road. The virtual vehicle, that was projected through a Varjo XR-3 head mounted display, approached from the right at a speed of 30 km/h while 4 interfaces (2x world-locked, head-locked, and vehicle-locked) appeared to communicate the vehicle's intention towards the participants, in addition to a no-interface baseline. Participants were tasked with indicating when they were willing to cross through the push of a remote button from which their *Willingness to cross* and *Decision certainty* could be derived. Subjective data was collected after the trials and after the experiment through interviews and a questionnaire respectively.

Results: Results suggest a positive effect of the AV2P interfaces on the *Willingness to cross* and *Decision certainty*, although statistically not significant. In other words, *Willingness to cross* increases when the vehicle indicates that it will yield, and decreases when the vehicle communicates that it will not yield. *Decision certainty* also increases when an interface is present compared to the no-interface baseline. Moreover, participants indicated using the interfaces as a tool to validate their own decisions. Compared to the CAVE study, subjective intuitiveness ratings replicate in terms of observing higher intuitiveness of the interfaces than the no-interface baseline. However, the intuitiveness ratings were higher in the CAVE study than the real world AR experiment. Furthermore, the order of the top 3 most preferred interfaces ranking is in the opposite order. Both differences suggest that the increased ecological validity of the real world AR experiment introduces new insights into participants' perception of interfaces. The Van der Laan acceptance scale shows that participants believe interfaces to be useful and satisfying overall.

Conclusion: The experiments suggest that AV2P interfaces have a positive effect on the crossing behaviour of pedestrians. Furthermore, participants indicate using the interfaces as a tool to validate their own decision, which increases confidence in their decisions. Although results partially replicate a previous virtual environment study, there are differences that suggest that real world AR experiments provide valuable insights into participants' perception of interfaces in a more realistic experiment.

Table of contents

List of figures

List of Tables

1. Introduction

According to the United Nations (UN), over 300.000 pedestrian deaths occurred around the world in 2016 (United Nations, 2016 as cited in World Health Organization, 2018) accounting for 23% of all road deaths. Furthermore, pedestrians are nine times more at risk of death than the occupants of a car, per kilometre travelled (Global Mobility Report, 2017). In addition, more than half of all road injuries involve vulnerable road users (VRUs), such as cyclists and pedestrians (United Nations, 2016). In an effort to reduce this threat, the UN has included a target in the UN Sustainable Development Goals: halving the amount of traffic deaths by 2030 (United Nations, 2020). In this resolution (A/RES/74/299), the UN General Assembly notes “that continuous progress of automotive and digital technologies could improve road safety, including through the progressive development of highly and fully automated vehicles in road traffic” (United Nations, 2020, p4). Apart from this resolution, little is mentioned about specific innovative digital and automated vehicle technologies, such as connectivity for inter-vehicle communication, camera and sensor technology for assistive safety features, the use of (partially) automated vehicles (AVs), or vehicle-to-pedestrian (V2P) communication. Moreover, the US Department of Transportation concluded that 94% of all traffic incidents in 2015 were the result of human error (US Department of Transportation, 2017). Consequently, this suggests that the development of automated driving systems has the potential of reducing the amount of traffic incidents (Plumer, 2014, Reimer 2014).

Automated vehicles (AVs) are identified by various levels of self-driving capabilities (SAE International, 2021). Levels 0 to 2 describe driver support systems, examples of which include automatic emergency braking (level 0), adaptive cruise control (level 1), and adaptive cruise control combined with automatic lane centering (level 2). In all of these cases, the driver is considered to be in control, and is therefore expected to supervise and intervene to maintain safety. Vehicles with SAE level 3 or higher, are characterised by the car driving itself, and the driver only taking over when requested by the system (level 3) or not at all (levels 4 and 5). The difference between level 3 to 5 are the conditions in which the vehicle is expected to perform without human intervention, for instance automated driving in a traffic jam (level 3), a highly automated driverless taxi (level 4), or fully automated driving everywhere in all conditions (level 5). While replacing the driver with autonomous systems can potentially decrease the risk of human errors and improve overall traffic safety (Chou et al., 2016), this will remove the driver’s capability of explicitly communicating with other road users. For instance, De Winter et al. (2021) showed that in 25% of the cases in which pedestrians interact with a vehicle in a parking lot, the pedestrian’s gaze was aimed at the driver. Therefore, it has been suggested to include systems that can communicate with other road users explicitly to convey the AV’s plans and intentions when no active driver is present.

In fact, it has been shown that communicating an AV’s plans and intentions to surrounding road users could improve the interaction between AVs and other road users, resulting in increased trust (Habibovic et al., 2018; Omeiza et al., 2022; Riegler et al., 2021). Pedestrians’ willingness to cross increases when a vehicle’s intention to yield is communicated through an interface placed on the exterior of a vehicle (e.g.

Ackermans et al., 2020; Barendse, 2019; Bazilinskyy et al., 2019). Also, pedestrian confidence increases and a quicker response is observed when vehicles are fitted with an eHMI compared to no eHMI (Habibovic et al., 2018; Petzoldt et al., 2018; Wilbrink et al., 2021). However, when using interfaces placed on vehicles or embedded in infrastructure while dealing with several pedestrians, it can be difficult to discern to whom the communicated message applies (De Clercq et al., 2019; De Winter & Dodou, 2022; Tabone et al., 2021).

The development of extended reality (XR) presents an interesting opportunity for innovative solutions. XR is an overarching term to describe various immersive technologies, the most well known being virtual reality (VR) and augmented reality (AR) (Samini et al., 2021). VR is a technology in which a user is immersed in a completely virtual world, most commonly applied through a head mounted display (HMD), replacing the surrounding world entirely. AR on the other hand, allows for the overlay of virtual objects onto the real world, rather than replacing it entirely with a new digital environment (Azuma et al., 1997).

In the context of traffic interactions, AR technology can be used to overlay information onto the real world while being able to see the surrounding traffic environment. Wearable AR technology could enable users to receive V2P communication signals right in their field of view (FOV) such as navigation instructions or safety signals. Users could have the possibility of customising and personalising their AR interfaces to their own preference, eliminating language barriers, and potentially improving intuitiveness and clarity (Tabone, De Winter, et al., 2021). Furthermore, each user receives their own V2P communication message, eliminating the uncertainty of who is targeted by the interface. The AR interfaces could also be effective in addressing pedestrians who are distracted (e.g. by their AR headset). From a system implementation perspective, this could be an easier and cheaper method than installing hardware infrastructure in roads or onto vehicles (Tabone et al., 2021).

From a human factor research perspective, XR provides researchers with a great amount of experimental control and a safe method to conduct traffic experiments. Many previous studies researching the interaction between pedestrians and vehicles have therefore resorted to conducting experiments in virtual environments through HMDs (Ackermans et al., 2020; Colley et al., 2021; Deb et al., 2017; De Clercq et al., 2019; Hou et al., 2020; Mahadevan et al., 2019; Mok et al., 2022; Schmitt et al., 2022), immersive spaces such as the Cave Automatic Virtual Environment (CAVE) (Deb et al., 2017; Kaleefathullah et al., 2022; Schneider et al., 2021; Tabone et al., 2023b, or less immersive media such as video footage or photos (Tabone et al., 2023a; Pijnenburg, 2017; Kumaaravelu, 2022). However, conducting experiments in VR comes at the cost of ecological validity and realism (Kozlov & Johansen, 2010). For instance, participants can become distracted by the virtual environment (Tabone et al., 2023b), affecting their awareness during the experiment (Aromaa et al., 2020). Moreover, participants have indicated that the virtual experience influences their behaviour due to a decreased perception of danger compared to real world traffic scenarios (Mahadevan et al., 2019; Palmeiro et al., 2018). Consequently, researchers have suggested increasing ecological validity by bringing experiments closer to reality (e.g. Tabone et al., 2023a; Tabone et al., 2023b; Tran et al., 2022).

Therefore, the goal of this study is to bridge the gap of ecological validity that is present in research conducted in virtual environments, and build upon previous studies (Barendse, 2019, Tabone et al., 2023b) by conducting similar experiments in a real world environment. AR will be used to augment a virtual vehicle and display 4 different interfaces for AV2P communication and a no-interface baseline. This allows the experiment to be carried out with perfect repeatability, and without the risk of vehicle collisions with participants (Hasan, 2022), while retaining the realism of the surrounding real world (Maruhn et al., 2020). Furthermore, Aromaa et al. (2020) showed that participants using AR were aware of their real world surroundings, and that virtual objects did not cause substantial distractions. Admittedly, the experiment conducted by Aromaa et al. (2020) did not include a safety critical task. Still, conducting traffic experiments in AR can result in higher ecological validity compared to VR experiments (Matviienko et al., 2022), while simultaneously experimenting a method that more closely resembles a system in which pervasive and context aware AR becomes implemented for traffic signalling applications (Tabone et al., 2021).

The research question that this study aims to answer is:

What is the effect of AR AV2P interfaces on the willingness to cross and decision certainty of pedestrians, and to what extent does intuitiveness and preference ranking replicate from a virtual environment experiment?

This will be investigated by using a continuous objective measurement approach through the press of a button (Ackermans et al., 2020; Barendse, 2019; Peereboom et al., 2023). Several interfaces will be studied to determine their effects on pedestrian crossing behaviour and to compare to each other. This study will also investigate whether subjective intuitiveness ratings and interface rankings from an experiment conducted in a virtual CAVE environment (Tabone et al. 2023b) replicate in the real-world AR experiment. Additionally, participants' trust in and acceptance of an AR AV2P system will be researched.

In terms of objective measures, it is hypothesised that participants' *Willingness to cross* will more closely correspond to the vehicle's intentions when an interface is present versus the no-interface baseline. Therefore, participants are expected to show a greater *Willingness to cross* when the vehicle intends to yield, and show less *Willingness to cross* when the vehicle does not intend to yield. In addition, *Decision certainty* is expected to be higher when an interface is present than the no-interface baseline in both yielding and non-yielding conditions. In short this will replicate the general findings made by Barendse (2019). However, Barendse (2019) tested a relatively simple on-vehicle interface, which makes it difficult to compare interface specific performance. In terms of intuitiveness rating and the preferred interface ranking, the expectation is that the AR experiments will replicate the results from Tabone et al. (2023b), who concluded that pedestrians prefer interfaces that use traditional traffic signals, are easily visible, and allow observation of the vehicle and the interface simultaneously.

2. Methods

An experiment was conducted using AR in a real world environment to investigate the effect of the AR AV2P communication interfaces on the crossing behaviour of pedestrians, specifically *Willingness to cross* and *Decision certainty*. During the within-subject study, all participants were exposed to 4 interface conditions and a no-interface baseline, two vehicle behaviour conditions (yielding and non-yielding), and three attention attractor stimulus location conditions during a 45-minute AR experiment. Each participant signed an informed consent form (see Appendix A) before the experiment, and completed a pre-experiment and post-experiment questionnaire created with Qualtrics XM (Qualtrics, 2023) (see Appendix B) to investigate demographics and subjective ratings respectively. The experiment procedure was approved by the TU Delft Human Research Ethics Committee, approval no. 3054.

2.1 Participants

A total of 28 participants (23 identified as male and 5 identified as female), aged between 19 and 59 (M = 27, SD = 9) were recruited for the within-subject study. In terms of nationality, 26 indicated being Dutch, 1 Chinese, and 1 is unknown. All but one participant were used to right-hand traffic, and 22 participants mainly used a bicycle as transportation. 25% of participants indicated walking less than 15 minutes a day. One participant indicated that they are colour-blind (see Appendix B-c).

Participants were recruited through an opportunistic sampling approach, targeting individuals within personal networks, colleagues, and friends among other sources, aiming for voluntary participation in the experiment. One participant was recruited after presenting about the proposed experiment.

2.2 AR Interfaces

The choice of the specific interfaces that were used in the experiment is based on Tabone et al. (2023b) (from Tabone, Lee, et al. (2021)) in which participants were exposed to 9 different AR eHMIs that communicated the vehicle's intention, in a CAVE setup. The interfaces can be split into three categories: world-locked, meaning the interface was locked to a stationary point; head-locked, being locked in the participant's field of view remaining stationary relative to the participant's head position; and vehicle-locked, which is locked onto the vehicle. From these results, a subset of the best performing interface for each category was chosen in terms of intuitiveness rating to further evaluate. In addition, the *Fixed pedestrian lights* interface was added, which is an additional world-locked interface. The reason to include this interface was the high intuitiveness score, and to allow for the comparison of it to the *Pedestrian lights HUD*, both of which are similar in design but different in implementation. Interface specific descriptions are elaborated below. Appendix C provides additional design specifications.

Virtual fence: The *Virtual fence* interface consists of a zebra crossing projected on the road combined with vertical semi-translucent walls surrounding the zebra (Fig. 1). The entire interface is 3.00 m tall, 2.50 m wide and 7.50 m long. A semi-translucent gate positioned in front of the pedestrian opens in the yielding condition, and remains closed in the non-yielding condition. This equips the interface with redundancy, creating robustness (Wickens et al., 2011): the red or green state of the walls, and whether the gate opens or not. The projected zebra crossing part of the *Virtual fence* interface is not considered as an additional level of redundancy, since vehicles are obliged to yield when a pedestrian intends to cross at a (zebra) crosswalk in The Netherlands (Ackermans et al., 2020).

Even though the tunnel-vision nature of the *Virtual fence* has shown to give participants a safe feeling and very high level of clarity and visibility, participants have pointed out that the semi-translucent walls can partially occlude the view of the surrounding road and traffic situation (Tabone et al., 2023b). This is expected to show a pronounced effect in the proposed study compared to the earlier virtual experiments due to the higher ecological validity.

Figure 1: Virtual Fence & Zebra interface augmented at experiment location recorded from right eye view. Left: yielding (green) state; Right: non-yielding (red) state.

Fixed pedestrian lights: The *Fixed pedestrian lights* interface closely resembles existing pedestrian traffic lights, and is locked to the world, remaining stationary across the street from where the pedestrian is positioned. The interface is 2 m tall, consisting of a pole with a box on top that displays either a red icon of a standing pedestrian, or a green icon of a walking pedestrian, vertically placed above each other (Fig. 2).

The expectation is to see only slightly better performance compared to baseline, because the interface is only visible when directly looking at it and requires pedestrians to shift their focus away from the approaching vehicle. Compared to the CAVE study, this interface could be perceived as less clear, since it might not stand out as much in the real world environment through AR. Due to the familiar interface design it will likely be intuitive for participants. However, also due to this familiarity, participants might become confused when noticing that there is no traffic light directed at the vehicle, which normally is the case when encountering such a pedestrian traffic light in the real world.

Figure 2: Fixed Pedestrian traffic light interface augmented at experiment location from right eye view. Left: yielding (green) state; Right: non-yielding (red) state.

Pedestrian lights HUD: The *Pedestrian lights HUD* is similar to the *Fixed pedestrian lights* interface in terms of design. However, instead of being world-locked and stationary, the *Pedestrian lights HUD* is mapped to the user's head position, resulting in a head up display styled interface (Fig. 3). The traffic light box has a dimension of 0.20 m wide and 0.38 m tall. It is placed at 1 m distance of the participant's head and is placed off-centre by 30 cm upwards and to the right. The interface is tilted by 15° around the vertical y-axis to compensate for this offset and face the user. The signalling principle is identical to the *Fixed pedestrian lights*, using a red and green icon of a stationary and walking pedestrian respectively.

High performance, both objectively, and in terms of intuitiveness and clarity, is expected due to resemblance of current traffic signals and constant visibility in the user's field of view (FOV) allowing participants to observe both the interface and the approaching vehicle simultaneously. Additionally, the interface is also expected to perform better than the *Fixed pedestrian lights* in terms of willingness to cross and decision certainty.
Planes on vehicle: The *Planes on vehicle* is a vehicle-mapped interface consisting of

Figure 3: Pedestrian traffic light HUD interface augmented at experiment location from right eye view. Left: yielding (green) state; Right: non-yielding (red) state.

a red plane with an open stop-hand icon, or a green plane with an icon depicting a pedestrian crossing a zebra crosswalk (Fig. 4). The plane is locked to the vehicle's position and located above the engine bonnet. The plane has a dimension of 2.1 m wide and 1.58 m tall (4:3 aspect ratio) and is tilted by 54° to be positioned approximately parallel with the vehicle's windshield. Redundancy is provided by using a coloured plane and an icon.

The expectation is that some participants indicate that this interface blocks the view of the vehicle and of an expected driver, decreasing their perceived safety (Mok et al., 2022). Furthermore, the interface may require relatively high focused attention to understand due to the interface moving with the vehicle and changing in size. Participants have also previously indicated that it is difficult to discern whether the interface gives a command to the pedestrian or signals what the vehicle itself will do (Tabone et al., 2023a). On the other hand, participants can observe the interface and vehicle simultaneously because the interface is placed on the vehicle.

Figure 4: Planes on vehicle interface augmented at experiment location from right eye view. Left: yielding (green) state; Right: non-yielding (red) state.

The remaining interfaces that were investigated by Tabone et al. (2023b) (Augmented zebra crossing, Conspicuous looming planes, Field of safe travel, Nudge HUD, Phantom car) were not included in this study either due to similarities with other interfaces (e.g. Nudge HUD and Pedestrian lights HUD), or due to bad performance relative to the other interfaces for which it remains unlikely to perform well.

2.3 Experiment design

Participants performed the role of a pedestrian positioned on the side of an unmarked road. Their main objective was to indicate their willingness to cross through a remote button, which the participants pressed and held to indicate *Willingness to cross*, and released to signal unwillingness to cross. The use of a button, as opposed to a real crossing, might be perceived as a step backwards in ecological validity or realism. However, when considering the alternative of measuring crossing intention by means of a real crossing, several challenges were expected to occur. First of all, it is difficult to discern when a real crossing has precisely been initiated. There are many possibilities to define the exact start of a crossing. For instance, this can be when the participant has performed the first step forward, or when the participant has passed a certain point (Schneider et al., 2021). This can also vary amongst participants. Second, a real crossing can only be performed once per trial, after which the experiment would have to be reset and performed again, inhibiting the possibility of a continuous measurement of the participants' *Willingness to cross* during the entire trial. For instance, if the participant reacts very quickly to the AR interface and crosses early in the trial, there will be no data of the remainder of that trial, only yielding one snapshot for each trial. Therefore, the input through a button results in easy to measure binary data, and a continuous depiction of a participant's *Willingness to cross* during the entire trial. Third, the XR HMD was required to be tethered to the computer to receive power and exchange data. The participants' movements were therefore limited and a road crossing could not be performed. In the end, remaining stationary would be most practical in terms of hardware and effective in terms of measurements for the experiment.

An attention attractor stimulus in the form of a cyan circle ring (RGB: 0, 255, 255; see Appendix C) was added. The colour Cyan was chosen because it has been shown to be a neutral colour, having no meaning in modern day traffic (Bazilinskyy et al., 2020). The goal of this attention attractor stimulus was to enforce an initial distraction, replicating a pedestrian that has initially not seen the approaching vehicle, a method also applied by Tabone et al. (2023b). The attention attractor stimulus would appear in front, to the left or to the right of the participant. To enforce the replicated distraction, the vehicle would remain stationary after starting the trial until the participant had looked into a circle for 1 s. An additional benefit of this attention attractor stimulus was that participants' gaze at the start of each trial would be consistent.

Figure 5: Experiment location at TU Delft 3mE faculty. Left: Empty lot behind the faculty building with setup in front of door (the van is not part of the experiment); Right: Participant standing where the experiment is conducted.

2.3.1 Experiment location

An important aspect of ensuring high ecological validity was conducting the experiment in the real world, in which encountering the virtual vehicle would feel natural. Since the participant would be using augmented reality, an indoor location would not suffice. Hence, a closed off road would be the best option, since it is the exact environment in which this technology would be used in the future. However, there were several challenges with respect to the experiment setup. Mainly, the hardware required a connection to the computer in order to receive power. Second, the hardware would have to remain safe in case of bad weather, such as rain or heavy winds. Still, an outdoor space that is sufficiently large to accommodate the virtual experiment was necessary, as the virtual car needed enough space to approach the pedestrian realistically.

An empty lot was found behind the TU Delft Mechanical, Maritime and Material Engineering (3mE) faculty (see Fig. 5). This area was quiet as no traffic was allowed there. A great benefit of this location was the ability to place the computer inside a door and connect it to the Varjo XR-3 and the two SteamVR Base Stations 2.0 (HTC VIVE, 2019) which would be placed just outside of the door. This also prevented the need to setup and remove all of the gear every time it would be unattended, as most of the gear could remain close to the experiment location and indoors.

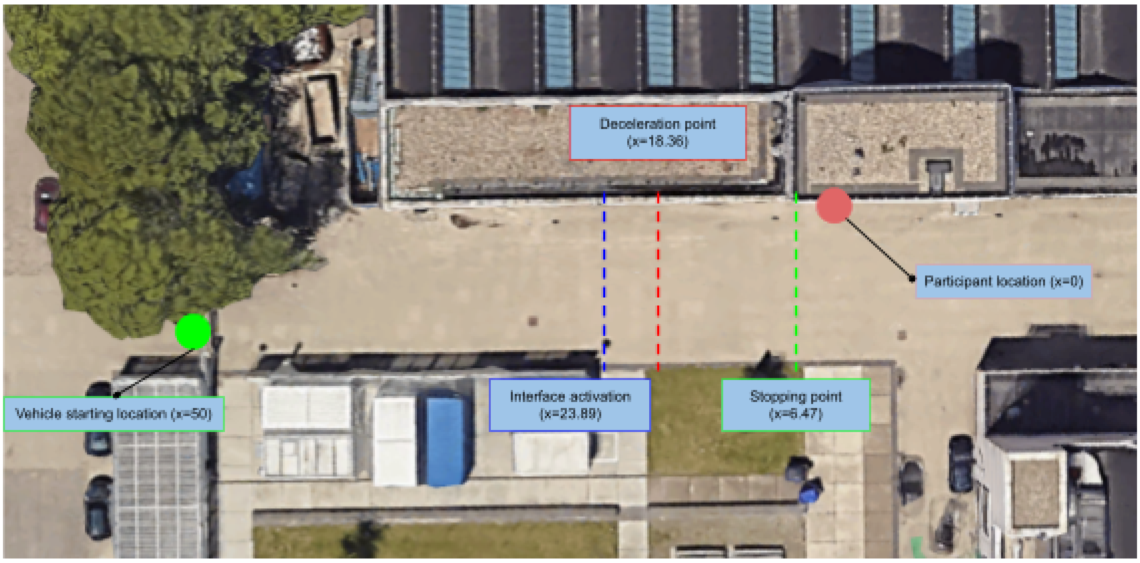

Figure 6: Experiment location with trigger locations at TU Delft 3mE faculty.
Source: Google Earth

2.3.2 Experiment simulations

The experiment simulations were designed, built and run on an Alienware with an Intel® Core™i7-9700K CPU @ 3.60 GHz, 16 GB RAM, and 16 GB Nvidia GeForce RTX 2080 Ti GPU with 11 GB of dedicated video memory. Unity 2021.3.13f1 (Unity, 2022) was used to build the AR simulation. To accomplish the vehicle's repeatable behaviour, three triggers (see Fig. 6) were placed and additional C# scripts established the vehicle's driving behaviour accordingly.

The Varjo XR-3 HMD was used in conjunction with two SteamVR Base Stations 2.0 (HTC VIVE, 2019) trackers to perform outside-in tracking (see Section 2.3.3 for more detail about tracking). A headphone was used for the virtual vehicle's audio, and was plugged into the Varjo XR-3 by means of an 3.5 mm auxiliary cable. The headphone did not include any active noise-cancelling, and no special passive-noise-cancelling, so surrounding real world environment noise could be heard.
In Unity, an invisible road was created by stretching a 10x10 m ground plane, resulting in a ground plane that was 1 km long and 10 m wide. This *Ground plane* was not visible for the participants and purely functioned as a rigid ground of the AR vehicle to drive on. The height of the *Main Camera* was set to 1.70 m. When using the Varjo XR-3

HMD during the experiment, this height corresponds to the distance between the participants' eyes and the *Ground planes*. A free car model asset was used and imported into Unity (Final Form Studio, 2021), representing a grey Audi sedan, measuring a length of 4.95 m, width of 2.10 m, and height of 1.35 m. The vehicle's wheel movements were animated to match its velocity. Audio was added to create a speed-sensitive engine sound and an idle engine sound, both with a doppler effect applied.

Initially, the intention was to position the virtual objects relative to a virtual origin, the *AR Origin*, and positioning the *AR Origin* in the real world using Varjo reference markers (Varjo Technologies, 2023). When testing outdoors however, the markers would reflect sunlight and would not be detected unless placed in very close proximity to the participant, which was not practical. The solution was to manually position the *AR Origin* according to the initial calibration position and orientation of the Varjo XR-3 HMD. That allowed AR assets (i.e. vehicle, ground plane, interfaces) to be placed relative to the *AR Origin* and automatically be positioned correctly relative to each other (see Appendix figure C). When this was established, the AR assets appeared to be too small in the real world. This was solved by increasing the *Scale* value of the *AR Origin,* ensuring all the connected assets scaled with the same value. The scale was checked by measuring the real world distance to the starting location where the virtual vehicle appeared to be located and comparing it to the virtual distance value that was assigned to the vehicle.

The interface locations were configured beforehand, either locked to the world, the vehicle, or the participants' head position. World-locked interfaces were mapped to the *AR Origin*, the vehicle-locked interface was mapped to the virtual vehicle asset (sedan-car-01), and the HUD interface was mapped to the headset position (XRRig; see Appendix Fig. C). The scripts activating the interfaces would instantiate the interface at these predefined locations. Appendix C provides more technical specifications about the interface design and positioning.

A. Vehicle behaviour

In the experiment conducted by Tabone et al. (2023b), the vehicle travelled at a speed of 30 mph (approx. 48 km/h, based on Kaleefathullah et al. (2020)). However, the current experiment location imposed an issue of using a velocity of 48 km/h. With the vehicle travelling at a velocity of 48 km/h, the vehicle would have to be located at 100 m from the participant to achieve the desired approach behaviour. However, a gate was located 50 m to the right of the participant, which would result in an unrealistic depiction of the virtual car floating through the gate. Rotating the experiment by 180° would also result in an undesired outcome, as the 3mE faculty building was located less than 50 m from the participants location, causing the same issue. The issue was solved by reducing the vehicle velocity to 30 km/h, and adjusting the trigger distances according to the time difference between the interfaces. A benefit of this speed is that it more accurately resembles city traffic in which pedestrians and vehicles interact.

At a velocity of 50 km/h (rounded up to become a common urban speed limit in The Netherlands), the time-difference (virtual time to collision or TTC) from the starting location to the *interface activation* trigger is 3.48 s (see Fig. 6); from the *interface activation* trigger to the *deceleration point*: 0.74 s; and from the *deceleration point* to the final *stopping point*: 3.96 s. These TTCs were used to convert the distances of the triggers from the initially planned velocity of 50 km/h to a velocity of 30 km/h. In the end, there is an identical amount of time between each trigger, with the only difference being the velocity at which the vehicle travels. As a result, the virtual vehicle could be spawned at a distance of 50 m, from where it would start to drive. At 23.89 m, the first ***interface activation*** trigger was positioned, triggering the AR interface. At a distance of 18.36 m, the next ***deceleration point*** trigger was positioned. In the case of the yielding condition, this triggers the deceleration of the car. At 6.47 m, the ***stopping point*** trigger was placed, which triggers hard braking in the case of the yield condition. This trigger is created to establish a constant final stopping location. In the non-yielding condition, the deceleration point and stopping point triggers did not affect the vehicle's behaviour. Similarly, the interface activation trigger was disabled for the baseline condition.

B. Scripts

In order to run the trials, a script was created that allowed the experimenters to control the experiment conditions from within Unity using check boxes, and to number log files by participant number and trial number through an open text-box (see appendix D). The check boxes were used to select the yielding behaviour (yield or no-yield) and select which interface would be active.

Large language model ChatGPT (OpenAI, 2023) was used to aid the process of writing these scripts. See Appendix E for a step-by-step data processing procedure. This procedure was explained and written in several prompts that were inserted into ChatGPT. The large language model would return a suggested script which was then tested, often requiring minor adjustments to enable proper functioning.

Many scripts were created to control all the aspects of the simulation and experiment: vehicle behaviour, countdown timer, *Virtual fence* gate controller, interface activation (one for each interface), data logging, Varjo AR video passthrough, and the wheel animations.
The scripts are available via: https://github.com/thomasaleva/AV2P-Thesis2023

C. Augmented reality design

Unfortunately, the previously designed interfaces (Tabone et al., 2023b) could not be copied and pasted into the new Unity project. The main issue was that many interfaces depended on other files and scripts that were not properly transferred to the new project. Certain parts of the assets could be re-used from the repository pertaining to Tabone et al. (2021b). Therefore, it was concluded that the easiest way to proceed would be to recreate the interfaces, utilising the usable parts of the existing assets. Appendix C provides more technical information about the design of the interface.

2.3.3 AR hardware

Varjo XR-3 was used in the experiments as an augmented reality head mounted display (HMD) (*Varjo*, 2020). During the experiments, the HMD was used to augment the virtual vehicle and AR interfaces while participants could see the surrounding real world. Varjo XR-3 is a video pass-through XR HMD and has inside-out (tracking the outside world from the headset) or outside-in (tracking the headset's movements from the outside) tracking capabilities to accurately track participants' head movement. The headset had to be tethered to power and the computer, which limited the movement of the participants. The cables were managed using a fishing-rod-like pole that guided the extension cables from the PC to the participant's location.

In order to use Varjo in conjunction with Unity, the Varjo XR Plugin was installed. By default the Varjo XR-3 is not in video pass-through mode when the simulation is started in Unity, which effectively results in a VR experience. In this case, video pass-through was activated through a script when the simulation was started.

As mentioned before, Varjo XR-3 supports inside-out tracking. This works best in an indoor environment in which a user is surrounded by many objects that can be tracked. Varjo reference markers (Varjo Technologies, 2023) can be placed around the user to increase the performance of the tracking. Unfortunately this worked very poorly when tested for the outdoor setup, due to a lack of surrounding objects, and the reference markers not being recognized caused by sunlight reflections and because they were positioned too far. Therefore, outside-in tracking through two SteamVR Base Stations 2.0 (HTC VIVE, 2019) was chosen as the best method to perform user head tracking. This tracking system works by placing Base Stations around the area in which the Varjo XR-3 is used (see Fig. 7). The base stations emit infrared light which the Varjo XR-3 uses to determine its location and rotation. In the experiment, two base stations were used. Every time the setup was placed, the Varjo and Base Station had to be calibrated. The tracking performed quite well, although sometimes the tracking glitched, especially during experiments held in very light rain. Direct sunlight did not influence the tracking performance.

Figure 7: Participant wearing the tethered Varjo XR-3 HMD. One SteamVR base station can be seen on a tripod in the left of the image.

The Varjo XR-3 was also used to record eye-tracking. To enable eye-tracking, the participants' eyes were calibrated for high accuracy, which is different from the default eye-tracking that is requested when a participant wears the Varjo for the first time. The high accuracy calibration uses a process with multiple calibration locations around the participant's FOV to increase tracking accuracy. The video pass-through was recorded while a gaze dot indicated where the participant was looking. In addition, a log file was created to record the eye-tracking data. The eye-tracking data was recorded as secondary data, as it was very easy to record. However, the eye-tracking will not be used in this study as it is not the scope of this research.

A logitech R400 remote presenter was used as the remote control, on which the right arrow button was used as the button to indicate willingness to cross. This button also includes a small extruded dot on top, which is why this button was chosen. The receiver was placed in a USB-port in the Alienware desktop.

2.3.4 Experiment procedure

The experiment had three independent variables: Interface condition, the attention attractor stimulus location, and yielding behaviour. The interface conditions consisted of four interfaces and the baseline condition, resulting in a total of 5 blocks. The attention attractor stimulus location could be left, right or centre. Each of these conditions would be performed for both the yielding and non-yielding states of the vehicle's behaviour. Therefore, 3 (locations) * 2 (yielding states) = 6 trials were conducted within each block. This resulted in a total of 30 trials per participant. The order of the blocks, attention attractor locations and yielding behaviour states were randomised using the Latin Square method (Altman, 2013).

After signing a consent form (Appendix A), each participant was instructed to complete a pre-experiment questionnaire (Appendix B-a) to gather demographic information. The pre-experiment questionnaire also included the Affinity for Technology (ATI) scale (Franke et al., 2018), followed by inquiring whether participants have experienced VR or AR technology before. Subsequently, the participant was instructed to read a written experiment explanation document (see Appendix F). Next, the Varjo XR-3 HMD was introduced to the participants and another oral explanation of the experiment was repeated. Subsequently, participants put on the Varjo XR-3 HMD, were handed the remote button, and initial eye-calibration would start automatically. Each participant required 45-60 minutes to complete the entire experiment. Some participants completed the pre-experiment questionnaire before their scheduled experiment slot, which saved time. The time spent wearing the AR HMD amounted to approximately 30 minutes.

When the trial started, the virtual vehicle would spawn at its starting location, and the attention attractor stimulus appeared in front, to the left, or to the right of the participant at a distance of approximately 10 m. After looking into the circle for 1 s, the vehicle approached from the right side at a distance of 50 m travelling with a speed of 30 km/h. For the yielding condition, the vehicle would pass the interface activation trigger (Fig. 6), at which point the interface would become visible for the participant. Shortly after, the deceleration point would be passed, where the yielding vehicle would start to decelerate until the stopping point, where the vehicle would come to a standstill. For the non-yielding condition, the vehicle would also pass the interface activation trigger, showing the non-yielding state of the interface. The deceleration point and stopping point triggers did not affect the non-yielding vehicle, resulting in the vehicle continuing to drive at 30 km/h.

Exactly 10.0 s after the vehicle started driving, a statement appeared in front of the participant (Fig. 8). The participant was asked to indicate to what extent they agreed to a statement: “This interface/situation was intuitive for signalling “Please do cross the road”” for the yielding condition; or “This interface/situation was intuitive for signalling “Please do not cross the road”” for the non-yielding condition. Participants verbally rated to what extent they agreed with this statement using a 7-point Likert scale (Fully disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Fully agree). This moment also marked the end of the trial, at which the participants were no longer required to indicate their willingness to cross through the remote button. After each block of 6 trials, a series of open questions were asked about the experience regarding the interface design, the timing of the interface appearance, the decision making behaviour of the participant, preference between yielding or non-yielding state and their wellbeing regarding the AR HMD, according to the misery scale (MISC) (Bos, 2015), which was all recorded on a voice-recorder.

Directly after the experiment, participants completed the post-experiment questionnaire (see Appendix B-b). The questionnaire was divided into several sections relating to the AR experience (participants’ awareness, level of realism, and whether they behaved similarly to in the real world), preferred interface ranking, interface-specific questions regarding intuitiveness, and a final section about the acceptance of the AR AV2P communication technology. The latter section uses the Van der Laan acceptance scale (Van der Laan et al., 1997) which evaluates the level of usefulness and satisfaction by

Figure 8: Post-trial intuitiveness question and a stationary yielding vehicle.

using a five point scale [-2; 2] to assess the acceptance in two dimensions: usefulness and satisfaction. The questionnaire scales have been modified so the most positive outcome is indicated by a high score. All questionnaire sections use a 7 point scale except for the Van der Laan acceptance scale and adoption questions which use a 5 point scale. The interface ranking section did not use a scale, but rather asked participants to rank the interfaces from most to least favourite. For the interface ranking, the mean ranking score was determined per interface. Open questions were also added to allow participants to insert subjective answers, such as suggestions.

2.4 Dependent measures

The dependent measures of the objective data consist of the willingness to cross and the decision certainty. These measures are both based on the registered button presses from the remote control.

- **Willingness to cross:** *Willingness to cross* is determined by assessing the amount of registered button presses for each 0.5 m interval of vehicle distance across all participants and trials per interface condition. This amount is then divided by the total number of data points (including when the remote button is not pressed) under the same conditions, which then yields the *willingness to cross* percentage.
- **Decision certainty**: *Decision certainty* is determined by counting the number of decision reversals of the participants for each trial. This number is then averaged across each participant for each interface condition, and subsequently averaged across participants. A lower number indicates a higher certainty, thus suggesting high clarity, good comprehension, and trust in the interface (Barendse, 2019).

The trial data was logged in log-files, consisting of 11 columns: *Participant number; Trial number; Date & time; Runtime; Vehicle position X; Vehicle position Y; Vehicle position Z; Vehicle velocity; Vehicle acceleration; Button press (boolean); Circle nudge location*. This data was logged at 50 Hz into an Excel spreadsheet for each trial.

2.5 Data collection and analysis

All the data from the remote button and the vehicle data from Unity was collected in Excel sheet log files. All 28 participants resulted in a total of 839 log files (participant 1 trial 27 was an empty, perhaps corrupted log file), of which 420 for the yielding condition and 419 for the non-yielding condition.

During some trials, it was noticed that the Logitech R400 remote did not always connect to the remote button receiver placed in the PC. Participants often held their hands with the remote together in front of their body, blocking the direct line of sight from the remote towards the PC. Once this was noticed, participants were instructed before and regularly during the experiment to hold their remote next to their body, on the side located closest to the PC. Assuming all participants would press the button at least once during the yielding scenario, 67 log-files that did not record any button press in the yielding scenario were deleted from the data set. In the case of the non-yielding scenario, it could be the case that some participants did not press the button at all during the experiment, so no log files were deleted from the non-yielding data set.
Python was used to analyse the log file data in bulk. The log file data was cleaned and sorted for the yielding or non-yielding condition. ChatGPT was used in this phase to speed up the process of creating the scripts (OpenAI, 2023). The step by step method is described in Appendix E. This method was presented to ChatGPT in prompts describing all the sub steps. These prompts include an explanation of the log-file structure in terms of which columns hold certain data. Furthermore the prompts explained how the files were divided into folders per participant number, trial number, interface condition and yielding condition. All of these aspects were included in the ChatGPT prompts while asking to provide a script that could accomplish the data

handling, cleaning and visualisation. The scripts that ChatGPT returned were useful as a base, but often not completely functional yet, therefore requiring some adjustments before working properly.

Statistical analysis was performed using SPSS software and Python was used for visualisation purposes. The effect of the interface conditions on the participants' *Willingness to cross* was analysed by using the percentage of the trial distance that participants indicated being willing to cross. Subsequently a one-way repeated measures ANOVA test was performed, including the Bonferroni correction to determine significance. The same analysis was performed for *Decision certainty*.

3. Results

3.1 Willingness to cross

The *Willingness to cross* is shown in figure 9 as the percentage of participants that are willing to cross throughout the duration of the trial. In the plot, the vertical lines indicate the trigger locations: the black line indicates the *interface activation* distance, the red line indicates the *deceleration point*, where the vehicle undertakes starts decelerating in the yielding state, and the green line indicates the *stopping point* of the yielding vehicle.

When analysing the yielding condition, figure 9 shows that not all participants are willing to cross when the trial starts. Once the vehicle approach is noticed from the starting distance (50 m), the willingness increases, as participants start to understand the situation and feel that the vehicle is far enough to cross. Before the interface appears (*interface activation* trigger at 23.89 m), the *Willingness to cross* varies between 20% and 60%, mostly hovering between 40%-55% for the yielding condition, and 30%-45% for the non-yielding condition. For most interfaces, the *Willingness to cross* decreases from here as the vehicle approaches the participant and no interface is present yet.

As soon as the interface appears, the *Willingness to cross* surprisingly does not increase immediately in the yielding condition, but rather stops the declining trend and remains mostly constant. Only after the vehicle starts decelerating, does the upward trend in *Willingness to cross* become visible for the interface conditions. For the no-interface baseline condition however, *Willingness to cross* declines further. The *Pedestrian lights HUD* shows the lowest *Willingness to cross* after interface activation. This could be caused by some participants experiencing part of the *Pedestrian lights HUD* to be out of their field of view, due to not properly wearing the headset. Several pedestrians also indicated that it was difficult to distinguish the red and green state on both *Pedestrian lights* interfaces, though mostly applying to the *Fixed pedestrian lights*. *Planes on vehicle* shows the highest willingness to cross when the interface is present and the vehicle has started decelerating.

For the non-yielding condition, the *Willingness to cross* steadily decreases to a minimum when the vehicle is closest to the participant, in line with expectations. Before interface activation the *Willingness to cross* remains quite constant. When the interface appears, a declining trend in the *Willingness to cross* is observed. Again, *Pedestrian lights HUD* and *Fixed pedestrian lights* show the lowest level of *Willingness to cross* after interface activation. In this case though, that could suggest that the interface works effectively in convincing participants that they should not cross the road.

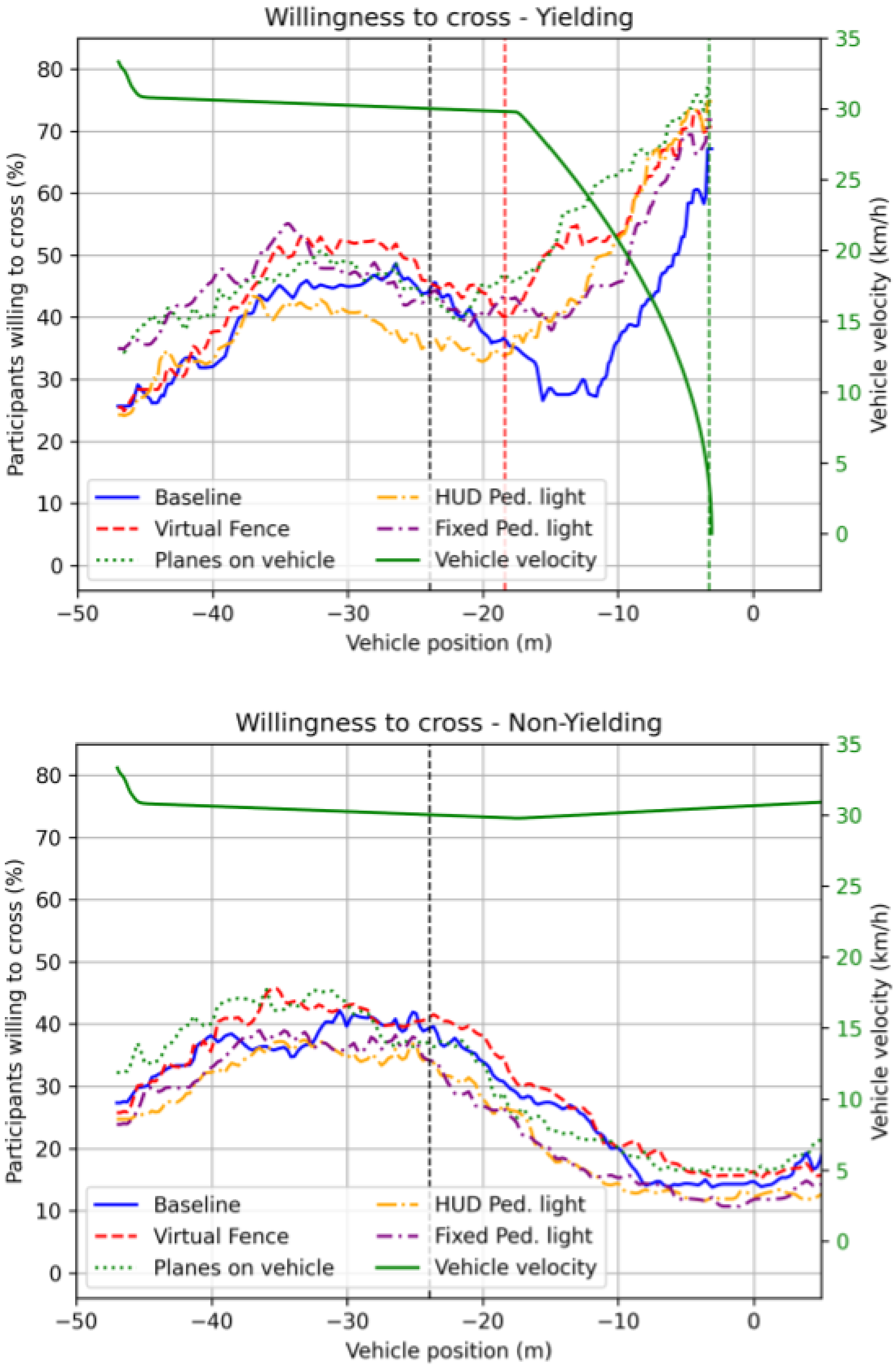

Figure 9: Willingness to cross - The percentage of participants indicating with the button that they are willing to cross during vehicle approach. The vertical lines represent the triggers at which the interface activated (black dotted), the vehicle decelerated (red dotted) and the vehicle stopped (green dotted).

Despite the vehicle communicating that it will not yield, participants show the highest *Willingness to cross* after interface activation for the *Virtual fence* interface, and show a delayed response to the *Planes on vehicle* interface. Although some participants mentioned “walking through a red light” in the post-block open questions (see Section 3.4.5), which could contribute to the high *Willingness to cross*, these observations suggest confusing and ambiguous interface designs. This could also explain why the *Willingness to cross* does not drop to zero when the vehicle passes the participants, which was unexpected.

For the yielding condition the one-way repeated-measures ANOVA test revealed that the interface conditions did not have a significant effect on the *Willingness to cross* compared to the baseline condition ($F(4, 88) = 1.024$, $p = .400$) for which Mauchly’s Test of Sphericity was assumed ($p = .276$). Post hoc pairwise comparisons using the Bonferroni correction indicated that there was an increase in mean *Willingness to cross* for all interface conditions compared to baseline. However, these differences were not statistically significant.

For the non-yielding scenario, Mauchly’s Test of Sphericity is confirmed ($p = .203$). Consequently the interface conditions also did not show a significant effect on the willingness to cross ($F(4, 108) = 0.353$, $p = .841$). Post hoc pairwise comparisons using the Bonferroni correction show a decrease in mean *Willingness to cross* for all interface conditions compared to baseline. However, these comparisons are not statistically significant. This indicates that the observed effects of the interfaces are not significant.

See Appendix G for details of the pairwise comparison.

3.2 Decision certainty

Decision certainty is defined by the amount of decision reversals during the trial. The more confident participants are of their crossing decision, the less decision reversals should be recorded (Barendse, 2019) (see Table 1). Overall, participants seem more certain of their decisions for the non-yielding vehicle, as the mean decision reversal rates are lower compared to the yielding condition. This aligns with post-block comments in which participants indicated that not crossing the road is their default decision when a vehicle is approaching, and that the interface is used as a tool to confirm participants’ decisions. Therefore, once the interface appears, it confirms the participant’s decision not to cross, leading to high certainty.

Furthermore, all interfaces improve *Decision certainty*. The *Planes on vehicle* interface results in the smallest amount of decision reversals, while the *Virtual fence* interface leads to the highest amount of decision reversals. Post-block interviews show that the *Virtual fence* interface could be interpreted as ambiguous, because it was not clear if the coloured walls targeted the pedestrian or the approaching vehicle. Meanwhile the simultaneous projection of a zebra crossing caused confusion in the non-yielding red state of the walls, as a zebra crossing motivated participants to cross the road.

Interface	**Mean (*SD*) reversals**	**N**	**Mean (*SD*) reversals**	**N**

	Yielding		Non-Yielding	
Baseline	2.60 (0.764)	25	2.00 (*0.938*)	28
Virtual fence	2.55 (*0.582*)	25	1.99 (*0.799*)	28
Fixed ped. lights	2.52 (*0.791*)	25	1.90 (*0.831*)	28
Ped. lights HUD	2.51 (*0.866*)	25	1.70 (*0.728*)	28
Planes on vehicle	2.41 (0.823)	25	1.86 (*0.767*)	28

Table 1: Mean number of button reversals per interface for the yielding and non-yielding condition.

For the yielding scenario, Mauchly's Test of Sphericity is rejected (p = .020). According to Greenhouse-Geisser, the interface conditions did not show a significant effect on the willingness to cross (F(2.817, 96) = 0.267, p = 0.837). Post hoc pairwise comparisons using the Bonferroni correction show that the decrease in the mean amount of decision reversals as a result of the interfaces is not statistically significant (p = 1.000).

For the non-yielding condition, Mauchly's Test of Sphericity can be assumed (p = 0.228). Again, the effect of the interface conditions on the *Decision certainty* does not show a significant effect (F(4, 108) = 1.345, p = .258). Post hoc pairwise comparison using the Bonferroni correction does show a decrease in the number of decision reversals throughout the trial. These decreases are statistically insignificant for all comparisons (p = 1.000, and p = 0.409 for the *Baseline-HUD Ped. Lights* comparison).

3.3 Post-trial intuitiveness

At the end of each trial the participants were asked to rate the intuitiveness of the interface. The mean values are shown for the yielding condition and the non-yielding condition in Table 2. Figure 10 shows a visualisation of the post-trial intuitiveness scores and provides a comparison to the post-trial intuitiveness scores reported by Tabone et al. (2023b) as a result of experiments in a CAVE environment. It stands out that the intuitiveness ratings of the CAVE study (Tabone et al. 2023b) all score higher than the intuitiveness ratings of the AR study, except for the baseline yielding state. The *Planes on vehicle* interface however, shows very similar results in both experiment environments.

Overall, intuitiveness scores of the yielding conditions are all higher than the non-yielding conditions per interface. Participants explained to have already made their choice not to cross before the interface appeared in the non-yielding condition, adding that not crossing is their default choice. Therefore the interface added less value to them, leading to a lower score for the non-yielding conditions. This shows that implicit communication remains an important factor in the road crossing decision making process of pedestrians. It also suggests a poor understanding of and adequate response to the question.

Interface	**Mean (*SD*)**	**N**	**Mean (*SD*)**	**N**

intuitiveness	Yielding		Non-Yielding	
Baseline	4.63 (*1.78*)	28	4.55 (*1.99*)	28
Virtual fence	6.02 (*0.877*)	28	5.31 (*1.15*)	28
Fixed ped. lights	5.48 (*1.23*)	28	4.67 (*1.64*)	28
Ped. lights HUD	5.96 (*1.05*)	28	5.58 (*1.46*)	28
Planes on vehicle	5.93 (*1.72*)	28	5.45 (*1.60*)	28

Table 2: Mean post-trial intuitiveness score per interface condition. 7-point scale

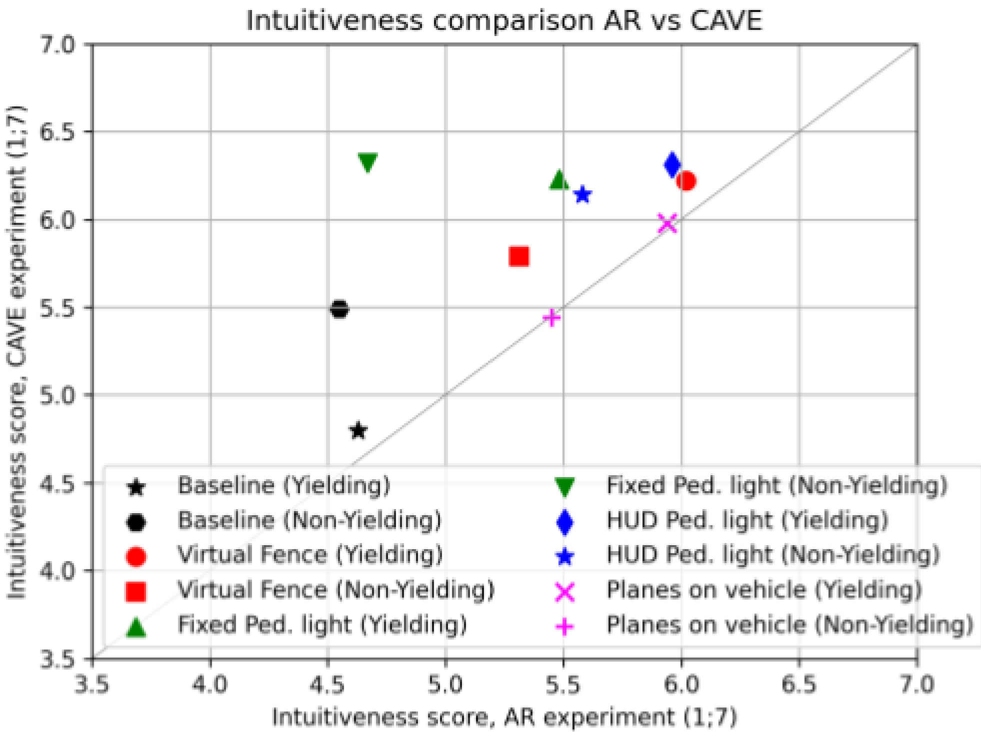

Figure 10: Scatter plot of mean post-trial intuitiveness scores from AR experiment. CAVE experiment (Tabone et al., 2023b).

3.4 Post-experiment Questionnaires

3.4.1 AR experience

Overall, most questions were answered to the positive side of the 7-point scale, indicating positive outcomes for the experiment. In terms of awareness (Q1-2), presence (Q3-5), and attention (Q6), the mean scores all indicated higher than 5 (see Appendix Table B-d). Questions comparing augmented reality to the real world, photos, videos, a virtual world or game (Q7-11) scored slightly higher than a neutral score of 4. This suggests that participants perceived the AR experience to be realistic, rather than having the feeling of being in a virtual world, game, or watching photos or videos, although only by a little. This is a positive outcome for the experiment, as the goal was to increase ecological validity and realism compared to VR experiments.

Regarding the perception of augmented objects and assets (i.e the virtual vehicle) (Q12-14) participants scored only slightly above neutral, implying that the augmented objects were perceived as slightly realistic. Participants also responded that they were captivated by the augmented assets (Q15), which indicates a distracting effect during the experiment and could affect participant behaviour. On the other hand, participants reported having the feeling that they behaved similarly to how they would behave in the real world (Q16) (see Appendix Table B-d).

3.4.2 Interface ranking

The most preferred interface was the *Pedestrian lights HUD*, followed closely by *Planes on Vehicle* and *Virtual Fence* (see Table 3). The *Fixed pedestrian lights* and baseline condition ranked as least preferred interfaces. This corresponds to Tabone et al. (2023b) who concluded that HUD interfaces were most preferred due to their good and continuous visibility. However, Tabone et al. (2023b) showed that the highest ranking interface of the CAVE study was the *Virtual fence* followed by *Planes on vehicle*, *Pedestrian lights HUD*, and *Fixed pedestrian lights*, not regarding the other interfaces that were tested in the CAVE study. This order of ranking does differ from the findings in the real world AR experiment, as the relative ranking of the top three interfaces is reversed. *Fixed pedestrian lights* scores lowest in both situations.

Interface	**Mean rank (*SD*)**	**N**
Baseline	4.00 (*1.28*)	28
Virtual fence	2.61 (*1.45*)	28
Fixed ped. lights	3.36 (*1.25*)	28
Ped. lights HUD	2.50 (*1.11*)	28
Planes on vehicle	2.54 (*1.43*)	28

Table 3: Mean interface preference rank. Ranks ranged from 1 (most preferred) to 5 (least preferred).

3.4.3 Van der Laan usefulness and satisfaction ratings

Figure 11 shows the mean scores for usefulness and satisfaction per interface and a semi-transparent scatter of the individual data points per interface. The interfaces all score in the upper right quadrant of the plot. This shows that participants perceive the AV2P communication interfaces as useful and satisfying. It is noticeable that the *Fixed pedestrian lights* interface scores worst of all interfaces, and the *Pedestrian lights HUD* interface scores closest to the top right of the graph, which corresponds to the interface ranking order. It can also be seen that there are several outliers in the bottom left quadrant of the graph, mostly consisting of the *Fixed pedestrian lights* and *Virtual fence* interfaces.

It appears that *Planes on vehicle* scores highest in terms of satisfaction, but second highest on the usefulness scale, close to *Virtual Fence* (see Table 4). Participants have indicated in the post-block questions that the red state of the *Planes on vehicle interface*, specifically the hand icon, caused confusion. In some cases, participants perceived the red state of the interface to indicate that the vehicle would stop.

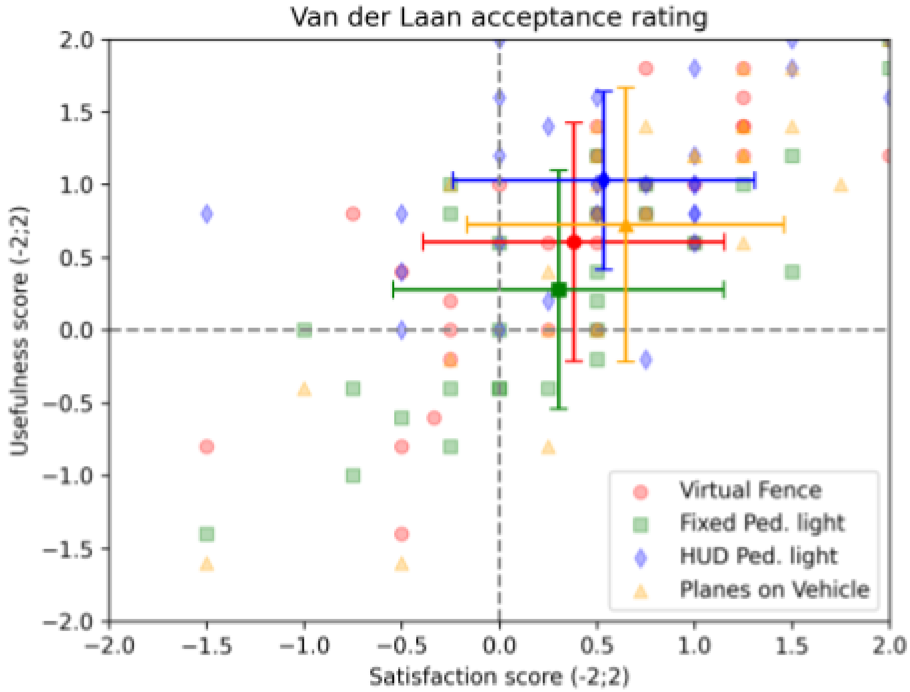

Figure 11: 2D error bars & scatter of Van der Laan acceptance scores

Interface	**Mean rank (*SD*)** Usefulness	**Mean rank (*SD*)** Satisfaction	**N**
Virtual fence	0.607 (*.834*)	0.381 (*.786*)	28
Fixed ped. lights	0.279 (*.833*)	0.304 (*.862*)	28
Ped. lights HUD	1.03 (*.625*)	0.536 (*.788*)	28
Planes on vehicle	0.726 (*.961*)	0.648 (*.827*)	28

Table 4: Mean Van der Laan acceptance scores: usefulness & satisfaction

3.4.4 Trust & acceptance

In this part of the questionnaire, participants were asked whether they trust the AV2P communication system through AR interfaces, and whether they would use such a system as an aid for road crossing. Participants were asked to rate the amount of trust they have per interface on a 7 point scale (see Table 5). Another column is added to convert the 7 point scale into a 5 point scale, to more easily compare it to the trust and adoption answers (see Table 6).

The results show that the *Pedestrian lights HUD* scores highest in trust, followed closely by *Virtual fence* and *Planes of vehicle*. Several participants indicated in the post-block interviews that some interfaces, such as the *Pedestrian lights HUD* and *Fixed pedestrian* *lights*, feel somewhat “disconnected” from the vehicle, therefore not feeling confident that the vehicle’s behaviour would correspond to what the interface communicated. When asked to compare it to other interfaces, participants felt the least amount of this “disconnect” with the *Planes on vehicle* interface, because it felt the vehicle itself communicated directly with them. Still, the results show that the *Pedestrian lights HUD* earned the highest amount of trust. This aligns with the stimulus-response compatibility issues pointed out by Tabone, De Winter, et al. (2021).

Questions regarding trust in the AV2P communication system in general (Table 6), and trusting the interfaces to make a crossing decision show similar results. Although, the general trust is skewed slightly higher than the interface-specific trust when comparing the converted 5 point scale of Table 5 to Table 6.

Mean trust per interface	**Mean (*SD*) 7-point scale**	**Mean (*SD*) converted to 5-point scale**	**N**
Virtual Fence	4.79 (*1.71*)	3.42 (*1.22*)	28
Fixed pedestrian lights	4.39 (*1.91*)	3.14 (*1.36*)	28
Pedestrian lights HUD	5.00 (*1.16*)	3.57 (*.829*)	28
Planes on vehicle	4.59 (*1.78*)	3.28 (*1.27*)	28

Table 5: Trust to base crossing decision on interface

Trust and adoption	**Mean (*SD*) 5-point scale**	**N**
I trust the AR interface to help me make my decision about when to cross	3.83 (*1.74*)	28
In general I trust such an AR system to communicate the correct information	3.75 (*1.26*)	28
If I would own an AR headset, I would be willing to use the AR interfaces as an aid for road crossing	3.17 (1.*27*)	28

Table 6: Trust & adoption of AR AV2P communication systems

In the hypothetical case that participants own an AR headset, the willingness to use AR interfaces as an aid for road crossing shows a response just above neutral. This neutral score is contrary to what the pre-experiment questionnaire suggests in which 89% of participants responded to use technical systems to their full extent. This could be attributed to the level of discomfort participants experienced from wearing the AR tethered headset during the experiment. Additionally, participants might have difficulty imagining the idea of regularly wearing an AR HMD for extended periods of time altogether. Unfortunately this was not specifically asked.

During the experiment, participants were asked to state their level of misery using the MISC scale (Bos, 2015). Across all experiments, 20% of participants reported a level of more than 0 at least once during the experiment. The MISC levels remained sufficiently low that the experiment could continue. The MISC scale only determines dizziness and nausea rather than physical discomfort of the AR headset. In fact, some participants indicated that the AR headset was not comfortable and quite heavy, when inquiring about their MISC rating.

3.4.5 Post-block open questions

Directly after each block of 6 experiment trials, participants were asked several open questions to comment on the interface and their experience with it. A summary of various recurring comments about general topics is shown below, and interface specific comments are summarised in Appendix H. These open comments were used to provide additional insights and explanations about participants' behaviour, which has been referred to on several occasions in the previous sections of this report.

Learning curve

- I used the first run of each interface block to get used to the new interface as I did not know what to expect.
- During the first trial of each block I had to figure out what I would see and what the interface meant, instead of representing a real life experience.

Timing

- Interfaces appeared late, so you had to decide quickly after the appearance.
- I did not know where to look to find the interface except for the HUD that was stuck in my field of view. I also had to get used to the new designs of the interfaces, as opposed to a traffic light which I am used to.
- Interfaces should appear earlier.
- Timing and positioning of interfaces could be optimised.

AR experience

- Combination of the real environment and VR feels fine.
- I do not remember hearing any audio.
- Might act differently in other (red. Real life) surroundings
- Didn't notice whether there was a driver in the virtual car or not, but in the real world I would unconsciously focus on making eye contact with the driver.
- The biggest disconnect between the AR environment and the real world is the lack of peripheral vision while wearing the headset. The car was flying a bit at the beginning and end of every trial.

Several participants also suggested adding multiple vehicles, including a vehicle that approaches from the left. Additionally, fully randomising the order of the interfaces was suggested. A recurring design improvement was showing a default (or non-yielding) state of the interface, and switching it to yielding or non-yielding when necessary. In that case, pedestrians would know where to look and therefore might show a shorter learning curve.

4. Discussion

The aim of this research was to investigate the effect of autonomous vehicle to pedestrian (AV2P) communication interfaces on pedestrians' road crossing behaviour in terms of *Willingness to cross* and *Decision certainty* in a real world AR experiment to increase ecological validity, which is a common limitation observed in previous experiments conducted in virtual environments. In order to increase the ecological validity of the experiments, experiments were conducted by projecting a virtual vehicle in the real world through an augmented reality HMD. Additionally, it was investigated whether interface intuitiveness ratings and interface preference ranking would replicate compared to a virtual environment study by Tabone et al. (2023b).

4.1 Pedestrian road crossing behaviour

Overall, the results show that participants' *Willingness to cross* increases for yielding vehicles communicating their intention to yield, and decreases for non-yielding vehicles communicating their intention not to yield, compared to the no-interface baseline. These findings align with previous studies (e.g. Barendse, 2019; Bazilinskyy et al., 2019; De Clercq et al., 2019; Deb et al., 2019; Hou et al., 2020; Mahadevan 2019). The effect of the interfaces on pedestrians' *Willingness to cross* is most noticeable in the yielding scenario, for which the *Willingness to cross* increases earlier than the baseline condition during the vehicle approach, aligning with earlier findings (e.g. Ackermans et al., 2020). Contrary to Tabone et al. (2023b), participants showed an unexpected reaction to the Virtual fence interface in the non-yielding scenario showing that the ambiguous design of interfaces can lead to adverse effects. The ambiguity of the *Virtual fence* interface could be attributed to participants not clearly perceiving whether the colour of the walls was targeted to the pedestrian or the vehicle, or to the contradictory nature of the presence of a zebra crossing projected on the road while displaying the non-yielding red state of the walls. This suggests reassessing interface designs to solve ambiguity, for instance by placing a red cross on the virtual zebra in the non-yielding state of the *Virtual fence* interface.

When comparing interfaces with the no-interface baseline, results show that an interface results in higher *Decision certainty* amongst participants both in the yielding and non-yielding condition. It is noticeable that the non-yielding scenario shows higher levels of *Decision certainty* altogether. Although this was not expected according to Barendse (2019), this can perhaps be explained by participants mentioning in the post-block interviews that not crossing the road is their initial default decision when a vehicle is approaching. When the interface subsequently appears in the red state, it then confirms the participant's current decision.

Although the aforementioned effects of interfaces on *Willingness to cross* and *Decision certainty* can be observed from the experiments, none of the effects of the interface conditions on pedestrians' crossing behaviour was significant. This insignificance is contrary to Barendse (2019), who showed a significant effect of the interfaces on *Decision certainty*, and on *Willingness to cross* for the yielding condition. It is expected that a larger sample size will result in significant results.

In terms of interface usage, participants repeatedly mentioned using the interface as a tool to validate their own decisions, continuing to primarily use implicit cues such as vehicle velocity, acceleration and sound. This finding is confirmed by the *Willingness to cross* not directly increasing after the interface appears in the yielding condition, but rather increasing after the interface has appeared and the vehicle starts decelerating. The post-experiment questionnaire regarding trust in AV2P communication systems confirmed this finding as well. Namely, pedestrians showed only a slight level of trust in the interface to base their crossing decision on. In other words, participants show that they do not trust the system enough to fully base their crossing decision solely on the interface, and prefer to use it as a tool to validate their own decisions that are based on implicit cues. More research is necessary to investigate the evolution of trust as participants experience the AV2P communication system for a longer period, experience errors, and experience more complex traffic scenarios, perhaps requiring a long-term study. In addition, interface design optimization and customisation can potentially increase the level of trust. Van der Laan's acceptance scale that measures usability of satisfaction of a system, shows positive results supporting earlier experiments studying AV2P communication systems (e.g. Barendse, 2019).

4.2 Replication of subjective results

From figure 10 it stands out that the intuitiveness ratings from the CAVE experiments conducted by Tabone et al. (2023b) yield higher results than the ratings returned after the real world AR experiment while both experiments use a similar group of participants. The lower score of intuitiveness could be caused by the real environment resulting in a less "clinical" situation. The increased ecological validity could have presented a higher amount of natural distractions, noises, sunlight etc. This underlines the importance of testing interfaces in an experiment with high ecological validity in order to optimise interface designs to improve intuitiveness.

The preferred interface ranking suggests that pedestrians prefer an interface that is clearly visible without requiring a shift in gaze to see it, confirming findings by Tabone et al. (2023b). However, the order of the top 3 interfaces (*Pedestrian light HUD, Planes on vehicle, and Virtual fence,* resp.) are in reverse order compared to the ranking from Tabone et al. (2023b). Combined with the unexpected response to the non-yielding state of the *Virtual fence*, this indicates that participants perceive interfaces differently in the real world AR experiment compared to the CAVE study (Tabone et al., 2023b). A possibility could be that participants were more focused on the area in front of them in the CAVE environment as they would perform their crossing, whereas in the AR experiment, participants were perhaps more capable of being aware of their surroundings because they did not physically perform a crossing. Also, in contrast to Tabone et al. (2023b), participants commented that the *Pedestrian light HUD* interface blocked their view of the road and the approaching vehicle. This was never mentioned for *Planes on vehicle*, which was expected to block the view of where a driver is supposed to be.

In terms of realism, participants rated the AR experience to be consistent with a real-world experience, indicating they feel present, aware, and alert in the environment, and that they show similar behaviour to real world situations. This is a positive outcome for the study, as this was the goal of conducting the experiment with increased ecological validity. Participants did indicate that they were captivated by the AR objects (i.e. the virtual vehicle), which could be distracting to the participants, and suggests that more realistic AR objects could increase ecological validity even further.

5. Conclusion

The experiment conducted in this study has achieved an increased level of ecological validity by using augmented reality in a real-world environment to overcome the common limitation of ecological validity and realism observed in previous virtual environment studies.. This experiment method has shown that participants feel present, aware and alert in the real world environment while experiencing traffic scenarios through an AR HMD. Furthermore, participants indicated to exhibit realistic behaviour, in contrast to virtual environments in which participants mentioned feeling a lack of real danger and being distracted by the virtual environment.

This study has observed that AR AV2P communication interfaces can improve road crossing decision making for pedestrians in terms of *Willingness to cross* corresponding more closely to a vehicle's intentions, and improved *Decision certainty* compared to the baseline condition, although no significance was shown. It was observed that some interfaces (i.e. *Virtual fence*) could lead to adverse behaviour due to ambiguous design characteristics, increasing *Willingness to cross* when a vehicle does not yield, which contradicts a study conducted in a CAVE environment (Tabone et al., 2023b). Furthermore, the post-trial intuitiveness yielded lower results in the real world AR experiments compared to the CAVE study, which could be caused by the presence of real world distractions. This suggests that conducting human factor research with increased ecological validity yields meaningful insights, and in this case, specifically motivates the reassessment of the *Virtual fence* interface.

In conclusion, participants prefer interfaces that allow them to watch the approaching vehicle while being able to see the interface, such as the *Pedestrian lights HUD* and the *Planes on vehicle* interfaces. Moreover, participants have indicated using the interface as a tool to validate their own decision, not yet trusting the interface to fully base their crossing decision on. Further research is required to determine which interface designs work best, and if customisation and experience affect the performance of the interfaces. Furthermore, experiments should be conducted with increasingly complex traffic scenarios to improve ecological validity.

6. Limitations and recommendations

The focus of this study was to investigate the effects of AV2P communication interfaces on the crossing behaviour of pedestrians. The technical implementation of such an envisioned system was assumed, such as the inter-connectivity network between vehicle and headset that allows for the communication of the vehicle's intention. Moreover, the Varjo XR-3, which is considered to be the state-of-the-art AR HMD at the time, still presents technical challenges such as tracking performance, limited field of view, user discomfort, impracticality, weight, and the fact that it is not wireless. In short, wearable AR technology is expected to require a lot of development before becoming commonly used.

Although the outside-in tracking performed quite well, it would sometimes lose connection which resulted in a sudden rotation of all the AR assets. This would cause brief discomfort for participants, as it would cause confusion and imbalance. Another technical limitation was the robustness of the connection to the remote button. Many participants held the remote button in front of their body within two hands, which would cause a bad connection between the remote button and the receiver placed in the PC. Participants had to be reminded often to hold the remote next to their body to ensure that the button presses were properly registered. It is recommended to increase the robustness of this connection either through a cable or by placing the receiver in front of the participant. Additionally, due to the logging script running from within Unity, the remote button presses would only register when Unity was the active window on the PC. Workflow could be improved if the remote button could be recorded in the background as well.

The location was located outdoors to achieve the highest possible ecological validity. However, the weather dependency made the experiment logistics more challenging. It is recommended to include a tent under which the experiment can be conducted without degrading the aspect of realism too much, both in very sunny and rainy weather to increase participant comfort. Ecological validity can be improved by closing down a road and placing the participants on the curb of a real road. In this study, participants stated that the environment and vehicle speed (30 km/h) imply a pedestrian-friendly environment, affecting their decision-making behaviour. Moreover, the traffic scenario in this study was simple, only including a single vehicle from one direction. Several participants suggested adding vehicles from different directions. For instance, the *Planes on vehicle* interface would perhaps perform differently, because pedestrians would have to look in multiple directions to understand the traffic situation. It is also interesting to study the effect of experiencing errors in the system, such as a false yielding interface of vice-versa, similarly to what has been shown by Holländer et al. (2019). The location also included several practical challenges: numerous TU Delft employees and maintenance personnel (see white van in Fig. 5, left) would park their vehicles and walk around the area to enter and exit the building. In addition, construction was planned at the same time as the experiments. This caused the experiment location to be busier with vehicles, construction equipment and construction workers than usual. In the end the experiment location was not as controlled as initially expected in terms of random distractions, and sometimes led to people walking through the AR assets while the experiment was being performed.

During the experiment it was noticed that participants had difficulty understanding all the information they received prior to starting the experiment. They had to remember a lot of information, and participants seemed to become acquainted with the experiment only after experiencing a number of trials. This could be perceived by participants forgetting to look at the circle before the vehicle would start driving, or forgetting to press the button. To solve this issue and ensure consistent performance, it is recommended to include voice-instruction when starting and ending the trials, such as “look in the circle for one second”, “Press the button”, and “end of the experiment”. Furthermore, whether the participants should press the remote button when the trial started, was not consistently implemented. Therefore a critical gap or initial crossing decision became difficult to analyse. It is suggested to create a consistent procedure for the initial button press as well.

In terms of interfaces, only egocentric (from the pedestrian point of view) non-textual interfaces were tested in this study, as was the case in Tabone et al., (2023). Further research could determine whether other styles of interfaces result in better performance. The method of using an AR HMD allows for interface customisation, the effect of which is also interesting to explore.

To optimise the design and working principles of the interfaces, it is recommended to analyse the eye-tracking data of this study. Attention must be dedicated to standardising certain aspects of eHMI designs, to streamline user comprehension of interfaces instead of creating systems with varying working principles (Dey et al., 2020). The eye-tracking data could provide valuable information as to where pedestrians look, which can be used for interface design improvements.

References

Ackermans, S., Dey, D., Ruijten, P. a. M., Cuijpers, R. H., & Pfleging, B. (2020). *The Effects of Explicit Intention Communication, Conspicuous Sensors, and Pedestrian Attitude in Interactions with Automated Vehicles*. https://doi.org/10.1145/3313831.3376197

Altman, N. (2013). *The Latin Square Design | STAT 503*. PennState: Statistics Online Courses. Retrieved June 29, 2023, from https://online.stat.psu.edu/stat503/lesson/4/4.3#:~:text=The%20randomization%20procedure%20for%20assigning,squares%20of%20the%20specified%20size.

Apple. (2023, June 20). *The Home View on Apple Vision Pro* [Video]. Apple Newsroom. https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/

Aromaa, S., Väätänen, A., Aaltonen, I., Goriachev, V., Helin, K., & Karjalainen, J. (2020). Awareness of the real-world environment when using augmented reality head-mounted display. *Applied Ergonomics*, *88*, 103145. https://doi.org/10.1016/j.apergo.2020.103145

Azuma, R. (1997). A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments*, *6*(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Bansal, P., & Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. *Transportation Research Part A-policy and Practice*, *95*, 49–63. https://doi.org/10.1016/j.tra.2016.10.013

Barendse, M. (2019). *External Human-Machine Interfaces on Autonomous Vehicles: The Effects of Information Type on Pedestrian Crossing Decisions* [Masterscriptie]. TU Delft.

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'?* https://doi.org/10.1109/smc42975.2020.9282998

Bazilinskyy, P., Kooijman, L., Dodou, D., & De Winter, J. C. F. (2021). How should external human-machine interfaces behave? Examining the effects of colour, position, message, activation distance, vehicle yielding, and visual distraction among 1,434 participants. *Applied Ergonomics*, *95*, 103450. https://doi.org/10.1016/j.apergo.2021.103450

Bos, J. E. (2015). Less sickness with more motion and/or mental distraction. *Journal of Vestibular Research: Equilibrium and Orientation*, *25*(1), 23–33. https://doi.org/10.3233/ves-150541

Chou, F., Shladover, S. E., & Bansal, G. (2016). *Coordinated merge control based on V2V communication*. https://doi.org/10.1109/vnc.2016.7835933

Colley, M., Li, S., & Rukzio, E. (2021). *Increasing Pedestrian Safety Using External Communication of Autonomous Vehicles for Signalling Hazards*. https://doi.org/10.1145/3447526.3472024

De Clercq, K., Dietrich, A., Velasco, J. R., De Winter, J. C. F., & Happee, R. (2019). External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. *Human Factors*, *61*(8), 1353–1370. https://doi.org/10.1177/0018720819836343

De Winter, J. C. F., Bazilinskyy, P., Wesdorp, D., De Vlam, V., Hopmans, B., Visscher, J., & Dodou, D. (2021). How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study. *Ergonomics*, *64*(6), 793–805. https://doi.org/10.1080/00140139.2020.1862310

De Winter, J. C. F., & 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

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., & Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics*, *65*, 449–460. https://doi.org/10.1016/j.apergo.2017.03.007

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

Drechsler, M. F., Peintner, J., Reway, F., Seifert, G., Riener, A., & Huber, W. (2022). MiRE, A Mixed Reality Environment for Testing of Automated Driving Functions. *IEEE Transactions on Vehicular Technology*, *71*(4), 3443–3456. https://doi.org/10.1109/tvt.2022.3160353

Eisma, Y. B., Van Bergen, S., Ter Brake, S. M., Hensen, M. T. T., Tempelaar, W. J., & De Winter, J. C. F. (2019). External Human–Machine Interfaces: The Effect of Display Location on Crossing Intentions and Eye Movements. *Information*, *11*(1), 13. https://doi.org/10.3390/info11010013

Final Form Studio. (2021, August 15). Sedan Car - 01 | 3D Land | Unity Asset Store. https://assetstore.unity.com/packages/3d/vehicles/land/sedan-car-01-190629

Franke, T., Attig, C., & Wessel, D. (2018). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Global Mobility Report. (2017). Sum4all. Retrieved July 8, 2023, from https://www.sum4all.org/publications/global-mobility-report-2017

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S., Saluäär, D., & Larsson, P. (2018). Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01336

Haddington, P., & Rauniomaa, M. (2013). Interaction Between Road Users. *Space and Culture*, *17*(2), 176–190. https://doi.org/10.1177/1206331213508498

Hasan, R. (2022). Pedestrian safety using the Internet of Things and sensors: Issues, challenges, and open problems. *Future Generation Computer Systems*, *134*, 187–203. https://doi.org/10.1016/j.future.2022.03.036

Holländer, K., Wintersberger, P., & Butz, A. (2019). *Overtrust in External Cues of Automated Vehicles*. https://doi.org/10.1145/3342197.3344528

Hong, J. Y., He, J., Lam, B., Gupta, R., & Gan, W. (2017). Spatial Audio for Soundscape Design: Recording and Reproduction. *Applied Sciences*, *7*(6), 627. https://doi.org/10.3390/app7060627

Hou, M., Mahadevan, K., Somanath, S., Sharlin, E., & Oehlberg, L. (2020). *Autonomous Vehicle-Cyclist Interaction: Peril and Promise*. https://doi.org/10.1145/3313831.3376884

HTC VIVE. (2019, May 1). *SteamVR Base Station 2.0 | VIVE European Union*. Retrieved June 29, 2023, from https://www.vive.com/eu/accessory/base-station2/

Ishihara, S. (1918). Tests for Color Blindness. *American Journal of Ophthalmology*, *1*(5), 376. https://doi.org/10.1016/s0002-9394(18)90663-x

Kaleefathullah, A. A., Merat, N., Lee, Y. H., Eisma, Y. B., Madigan, R., Garcia, J. A., & De Winter, J. C. F. (2020). External Human–Machine Interfaces Can Be Misleading: An Examination of Trust

Development and Misuse in a CAVE-Based Pedestrian Simulation Environment. *Human Factors*, *64*(6), 1070–1085. https://doi.org/10.1177/0018720820970751

Kozlov, M. D., & Johansen, M. K. (2010). Real Behavior in Virtual Environments: Psychology Experiments in a Simple Virtual-Reality Paradigm Using Video Games. *Cyberpsychology, Behavior, and Social Networking*, *13*(6), 711–714. https://doi.org/10.1089/cyber.2009.0310

Kumaaravelu, L. (2022). *Design of a real-time computer-vision and gaze-based pedestrian warning system* [Masterscriptie]. TU Delft.

Lee, Y. H., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C. J., Camara, F., Rothmueller, M., Vendelbo-Larsen, S. A., Rasmussen, P. D., Dietrich, A., Nathanael, D., Portouli, V., Schieben, A., & Merat, N. (2020). Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. *Cognition, Technology and Work*, *23*(2), 367–380. https://doi.org/10.1007/s10111-020-00635-y

Mahadevan, K., Sanoubari, E., Somanath, S., Young, J. B., & Sharlin, E. (2019). *AV-Pedestrian Interaction Design Using a Pedestrian Mixed Traffic Simulator*. https://doi.org/10.1145/3322276.3322328

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing Pedestrian Behavior in Augmented Reality — Proof of Concept. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. https://doi.org/10.1109/vr46266.2020.00051

Matviienko, A., Müller, F. H., Schön, D., Seesemann, P., Günther, S., & Mühlhäuser, M. (2022). BikeAR: Understanding Cyclists' Crossing Decision-Making at Uncontrolled Intersections using Augmented Reality. In *CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517560

Mercedes-Benz. (2015, April 28). *Mercedes-Benz F 015 Luxury in Motion: A Driving Experience of a Different Kind*. YouTube. Retrieved June 28, 2023, from https://www.youtube.com/watch?v=IWB4xj7EILg

Mok, C. S., Bazilinskyy, P., & De Winter, J. C. F. (2022). Stopping by looking: A driver-pedestrian interaction study in a coupled simulator using head-mounted displays with eye-tracking. *Applied Ergonomics*, *105*, 103825. https://doi.org/10.1016/j.apergo.2022.103825

Omeiza, D., Webb, H., Jirotka, M., & Kunze, L. (2022). Explanations in Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 10142–10162. https://doi.org/10.1109/tits.2021.3122865

OpenAI. (2023). ChatGPT.

Palmeiro, A., Van Der Kint, S., Vissers, L. S., Farah, H., De Winter, J. C. F., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. *Transportation Research Part F-traffic Psychology and Behaviour*, *58*, 1005–1020. https://doi.org/10.1016/j.trf.2018.07.020

Patella, S. M., Sportiello, S., Carrese, S., Bella, F., & Asdrubali, F. (2020). The Effect of a LED Lighting Crosswalk on Pedestrian Safety: Some Experimental Results. *Safety*, *6*(2), 20. https://doi.org/10.3390/safety6020020

Peereboom, J., Tabone, W., Dodou, D., & De Winter, J. (2023). Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for. . . *ResearchGate*. https://www.researchgate.net/publication/371509780_Head-locked_world-locked_or_conformal_diminished-reality_An_examination_of_different_AR_solutions_for_pedestrian_safety_in_occluded_scenarios

Penmetsa, P., Adanu, E. K., Wood, D., Wang, T., & Jones, S. J. (2019). Perceptions and expectations of autonomous vehicles – A snapshot of vulnerable road user opinion. *Technological Forecasting and Social Change*, *143*, 9–13. https://doi.org/10.1016/j.techfore.2019.02.010

Petzoldt, T., Schleinitz, K., & Banse, R. (2018). Potential safety effects of a frontal brake light for motor vehicles. *Iet Intelligent Transport Systems*, *12*(6), 449–453. https://doi.org/10.1049/iet-its.2017.0321

Pijnenburg, J. T. (2017). *Naturalism* [Masterscriptie]. TU Delft.

Plumer, B. (2014, May 28). 15 ways that self-driving cars could transform our lives. *Vox*. https://www.vox.com/2014/5/28/5756736/the-case-for-self-driving-cars

Qualtrics. (2023). *Qualtrics*. Retrieved June 30, 2023, from https://tudelft.eu.qualtrics.com/homepage/ui

Reimer, B. (2014). Driver Assistance Systems and the Transition to Automated Vehicles: A Path to Increase Older Adult Safety and Mobility? *Public Policy & Aging Report*, *24*(1), 27–31. https://doi.org/10.1093/ppar/prt006

Riegler, A., Riener, A., & Holzmann, C. (2021). A Systematic Review of Virtual Reality Applications for Automated Driving: 2009–2020. *Frontiers in Human Dynamics*, *3*. https://doi.org/10.3389/fhumd.2021.689856

SAE International. (2021, May 3). SAE Levels of Driving Automation™ Refined for Clarity and International Audience. https://www.sae.org/blog/sae-j3016-update

Samini, A., Palmerius, K. L., & Ljung, P. (2021). *A Review of Current, Complete Augmented Reality Solutions*. https://doi.org/10.1109/cw52790.2021.00015

Schmitt, P., Frazzoli, E., Coffey, A., Jeong, J., N, B., Clark, K. R., Khaw, A., Pham, L., Grigore, E. C., Ryan, K., Kothawade, S. S., Konopka, C., & Schmitt, C. A. (2022). Can Cars Gesture? A Case for Expressive Behavior Within Autonomous Vehicle and Pedestrian Interactions. *IEEE Robotics and Automation Letters*, *7*(2), 1416–1423. https://doi.org/10.1109/lra.2021.3138161

Schneemann, F., & Gohl, I. (2016). *Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments*. https://doi.org/10.1109/ivs.2016.7535361

Schneider, S., Maruhn, P., Dang, N., Pala, P. A., Cavallo, V., & Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and Head-Mounted Displays. *Human Factors*, *64*(7), 1210–1226. https://doi.org/10.1177/0018720820987446

SteamVR - Valve Corporation. (2023). SteamVR. Retrieved June 21, 2023, from https://www.steamvr.com/en/

Streater, J., Elias, J., Morrow, P. B., & Fiore, S. M. (2011). Towards an Interdisciplinary Understanding of Perspective for Human-Robot Teamwork. *Proceedings of the Human Factors and Ergonomics Society . . . Annual Meeting*. https://doi.org/10.1177/1071181311551308

Šucha, M., Dostál, D., & Risser, R. (2017). Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention*, *102*, 41–50. https://doi.org/10.1016/j.aap.2017.02.018

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. F., Lee, J. D., Martens, M., Merat, N., Norman, D. A., Sheridan, T. B., & Stanton, N. A. (2021). 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., Happee, R., García, J., Lee, Y. M., Lupetti, M. L., Merat, N., & De Winter, J. (2023a). 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., Happee, R., Yang, Y., & De Winter, J. (2023b). Immersive Insights: Evaluating Augmented Reality Interfaces for Pedestrians in a CAVE-Based Experiment. *ResearchGate*. https://www.researchgate.net/publication/370160064_Immersive_Insights_Evaluating_Augmented_Reality_Interfaces_for_Pedestrians_in_a_CAVE-Based_Experiment?_sg%5B0%5D=cuhj6eWQNE1D2SK2ipmn8fPDouHCd3ranuYTbHKt9rxMv3F0_cvrm46MHzqNiQgMsLx--_r1ZbjkDY5AAn1-t2bUVMDTQNBPpvJi5dy2.1QRXsTwjOfFLIURbCufhJfqLnFZyddErsoqppuoXJPjxEbsbSgssLIv48knKoS_sIjqVJ_AZ23FqA1jrv7NRqQ

Tabone, W., Lee, Y. H., Merat, N., Happee, R., & De Winter, J. C. F. (2021). *Towards future pedestrian-vehicle interactions: Introducing theoretically-supported AR prototypes*. https://doi.org/10.1145/3409118.3475149

Tran, T. T., 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*. https://doi.org/10.3389/fcomp.2022.866516

United Nations. (2016). *United Nations*. SDG Target 3.6 Road Traffic Injuries https://www.who.int/data/gho/data/themes/topics/sdg-target-3_6-road-traffic-injuries

United Nations. (2020, September 2). *United Nations A/RES/74/299* (A/RES/74/299). Resolution Adopted by the General Assembly on 31 August 2020. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/226/30/PDF/N2022630.pdf?OpenElement

Unity. (2022, November 4). *Unity*. Unity. https://unity.com/releases/editor/whats-new/2021.3.13

US Department of Transportation. (2017). *Automated Driving Systems 2.0 Voluntary Guidance*, DOT HS 812 442. https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C-emerging Technologies*, *5*(1), 1–10. https://doi.org/10.1016/s0968-090x(96)00025-3

Varjo. (2023, March 23). *Varjo XR-3 - The industry's highest resolution mixed reality headset* | *Varjo*. Varjo.com. https://varjo.com/products/xr-3/

Varjo Technologies. (2023, March 1). *Mixed reality – Varjo.com*. Varjo.com. https://varjo.com/use-center/get-to-know-your-headset/mixed-reality/#reference-marker

Wickens, C. D., Prinet, J. C., Hutchins, S., Sarter, N., & Sebok, A. (2011). Auditory-Visual Redundancy in Vehicle Control Interruptions: Two Meta-analyses. *Proceedings of the Human Factors and Ergonomics Society . . . Annual Meeting*, *55*(1), 1155–1159. https://doi.org/10.1177/1071181311551241

Wilbrink, M., Nuttelmann, M., & Oehl, M. (2021). *Scaling up Automated Vehicles' eHMI Communication Designs to Interactions with Multiple Pedestrians – Putting eHMIs to the Test*. https://doi.org/10.1145/3473682.3480277

World Health Organization. (2018). *Global status report on road safety 2018: summary*. https://apps.who.int/iris/handle/10665/277370

Yang, T., Kim, J., Fujiwara, H., Gil, H., Koo, J. H., & Kim, H. S. (2021). Recent Advances and Opportunities of Active Materials for Haptic Technologies in Virtual and Augmented Reality. *Advanced Functional Materials*, *31*(39), 2008831. https://doi.org/10.1002/adfm.202008831

Appendix

A. Consent form

Augmented-Reality Interfaces for Pedestrian-Vehicle Interaction

Consent Form for Participant

Researchers

MSc Student: Thomas Aleva (t.k.Aleva@student.tudelft.nl)
Supervisors: Wilbert Tabone, MSc (w.tabone@tudelft.nl), Dr. Dimitra Dodou (d.dodou@tudelft.nl), Prof.dr.ir. Joost C.F. de Winter (j.c.f.dewinter@tudelft.nl)

Location

Department of Cognitive Robotics
Faculty of Mechanical, Maritime and Material Engineering
Delft University of Technology
Mekelweg 2, 2628 CD, Delft
This document describes the purpose of this study, the experimental procedure, the right to withdraw and data handling. Read all sections carefully and answer the questions on page 2.

Aim of the research

The study aims to investigate pedestrians' crossing behaviour and self-reported preference for different Augmented Reality (AR) interfaces for pedestrian-vehicle interaction.

Experimental procedure

The study will be performed outdoors, in an area closed to traffic. You will be asked to stand at one side of a walking area while wearing an AR headset and holding a remote controller with a button. Via the headset, you will see virtual vehicles with various AR interfaces projected in the real environment. Your task is to hold a button on the controller when you feel safe to cross. You will also be asked to complete a questionnaire regarding your age and gender, experience with gaming and similar experiments, and your opinion about the interfaces. Participation will take about 45 minutes.

Risks

The use of an AR headset can cause different types of sickness: eyestrain, blurred vision, difficulty focusing, nausea, drowsiness, fatigue, or headache. These symptoms are similar to motion sickness. If you feel uncomfortable in any way, please inform the experiment so that the experiment is stopped.

Although the experiment is conducted outside, there is no accident risk, as the area is closed to traffic. The only vehicles you will be encountered will be virtual ones.

Right to withdraw

Your participation is completely voluntary, and you may stop at any time during the experiment for any reason. You have the right to refuse or withdraw from the experiment at any time, without negative consequences and without having to provide any explanation.

Data handling

All data collected during the experiment will be stored anonymously and will only be used for scientific research. You will not be personally identifiable in future publications based on this work, in data files shared with other researchers, or in data repositories. This consent form will be stored securely in a dedicated locker. Anonymized data will be published on a public repository.

Prevention of the spread of COVID-19

You may not participate if you show any symptoms indicative of COVID-19. You will be asked to disinfect your hands before touching any equipment.

Questions

If you have any questions, please contact one of the researchers.

Please respond to the following statements

Statement	Yes	No
I consent to participate voluntarily in this study.		
I have read and understood the information provided in this document.		
I adhere to the preventative measures with regard to COVID-19 explained above.		
I understand that I can withdraw from the study at any time without any negative consequences.		
I agree that the data collected during the experiment will be used for academic research and may be presented in a publication and public data repository.		

B. Questionnaires

a. Pre-experiment questionnaire

The questionnaire started with the participant number (Q0) which the experimenter would provide. Next, gender (Q1), age (Q2), nationality (Q3), the amount of time the participants had lived in The Netherlands (Q4), whether the participant was used to right-hand-traffic or left-hand-traffic (Q5), their occupation (Q6), and finally whether they use seeing aids (Q30). Next, the participant's openness to technology was asked (Q7) through a series of questions inquiring the participants about their interest in the details of systems, their willingness to understand technology, and how and why they use technology. The next section of the questionnaire related to AR and mobility, asking whether the participant had ever used VR (Q8) and AR (Q9) before, their daily walking time as a pedestrian (Q10), their primary mode of transportation (Q11), whether they possess a driving licence (Q12), and if so, when they obtained it (Q13), how often they drive each month (Q14), and finally how many kilometres they have driven in the past 12 months (Q15). The questionnaire ended with a colour blindness test (Ishihara, 1918, via Tabone et al., 2023b).

b. Post-experiment questionnaire

After the experiment, participants completed the post-experiment questionnaire about their experience with the AV and interfaces. Similar to the pre-experiment questionnaire, the post-experiment questionnaire started by asking the participant number (Q1) which was provided by the experimenter. The next section related to the AR experience by inquiring how aware participants were of their surroundings (Q2), how real the augmented objects seemed (Q3 & Q6), how consistent the AR experience felt with the real world (Q5), whether they felt present in the real world (Q7) and had a sense of “being there” (Q8). These questions were repeated in a series of sub-questions of Q4, including whether the participants had the feeling of playing a game, felt like perceiving photos or videos, and if they behaved in the same manner as they would have in a real world setting. Participants were then asked to rank the interface from most preferred to least preferred interface (Q11).

The next section was divided into interface specific blocks of questions. For each interface, the participant was shown a picture of the yielding state of the interface, as a reminder. Participants were asked to rate the intuitiveness and convincingness for the red and green state separately, and to what extent they trusted the interface to base their crossing decision on it (Q1 within the interface specific block). Subsequently participants were asked whether the interface was triggered too early or too late (Q2), too big or too small (Q3), and how clear (Q4) and visually attractive (Q5) the interface was. In the final section of the interface specific block, the Van der Laan acceptance scale was presented (Van Der Laan et al., 1997).

After the interface-specific question blocks, final section (Q60) related to the adoption of AR AV2P communication technology in the future. The participants were asked whether they thought the interfaces were useful for making a crossing decision (Q60.1), whether it helped them to make a decision (Q60.1), if they prefer interfaces to be mapped to the road (Q60.3) or mapped to their head movement (Q60.4), if participants would like the ability to customise the AR interfaces (Q60.5), if the system is generally trusted to provide communicate the correct information (Q60.6), and if participants would use AR interfaces as an aid for road crossing if they would own an AR headset (Q60.7).

c. General participant characteristics

- The 28 participants were aged between 19 and 57 (M = 27, SD = 9), of which 23 identified as male and 5 identified as female.
- In terms of nationality, 26 participants indicated being Dutch, 1 participant was Chinese and 1 participant did not respond to this question.
- One participant indicated having lived in The Netherlands for 1-5 years, while all other participants indicated living in The Netherlands for more than 10 years.
- One participant was used to left-hand-traffic while the remaining 27 participants were used to right-hand-traffic. None of the participants were used to both styles of traffic.
- Affinity for technology Interaction (ATI)
    - *I like to occupy myself in greater detail with technical systems*:
        - M4.89 SD1.14
    - *I like testing the functions of new technical systems*:
        - M4.89 SD1.21
    - *I predominantly deal with technical systems because I have to*:
        - M3.57 SD1.08
    - *When I have a new technical system in front of me, I try it out intensively*:
        - M4.36 SD 1.08
    - *I enjoy spending time becoming acquainted with a new technical system*:
        - M4.64 SD1.08
    - *I try to understand how a technical system works exactly*:
        - M4.36 SD1.20
    - *It is enough for me that a technical systems works; I don't care how or why*:
        - M2.68 SD1.17
    - *It is enough for me to know the basic functions of a technical system*:
        - M2.96 SD 1.09

- *I try to make full use of the capabilities of a technical system*:
    - M4.57 SD1.15
- 61% (*n=17*) participants used VR before and 32% (*n = 9*) participants had experienced AR before. 39% (*n = 11*) participants never used VR and 68% (*n = 19*) participants never used AR.
- Daily walking time
    - 25.00% (*n = 7*) <15 min
    - 42.86% (*n = 12*) 15-30 min
    - 28.57% (*n = 8*) 30-45 min
    - 3.57% (*n = 1*) >60 min
- Participants' primary mode of transportation was mainly the bicycle (79%; *n* = 22). 1 participant (3.6%) indicated it to be walking, 2 participants (7.1%) mainly used public transport and 3 participants (3.7%) mainly used their private vehicle.

d. AR Experience questionnaire

AR Experience Questionnaire	**Mean (*SD*)**	**N**
Aware of real world	5.21 (*1.95*)	28
Aware of surroundings	5.43 (*1.48*)	28
Had the sense of "being there"	5.79 (*.917*)	28
Did not feel present in surrounding environment*	5.46 (*1.32*)	28
Felt present in surrounding space	5.29 (*1.21*)	28
Paid attention to surroundings	5.75 (*.887*)	28
Sense of real world instead of virtual	5.21 (*1.37*)	28
Consistent with real world	4.96 (*.962*)	28
Felt like virtual world*	4.86 (*1.33*)	28
Felt like photo/video*	4.71 (*1.38*)	28
Felt Like game*	4.21 (*1.42*)	28
Realness AR Objects	4.46 (*1.29*)	28
How real did augmented objects seem	4.50 (*1.53*)	28
Virtual assets belonged environment	4.61 (*1.29*)	28
Captivated by virtual objects*	3.79 (*1.50*)	28
Behaved the same as in real-world situation	5.54 (*1.40*)	28

*Table B-d: AR Experience scores from post-experiment questionnaire. 7-point scale. *The scores have been adjusted so that a high score results in a positive outcome for the experiment.*

C. Interface designs

The RGB-colours were fully green (RGBa: 0, 255, 0, 125) for the green interface (Fig. B1), fully red (RGBa: 255, 0, 0, 180) for the red interfaces (Fig. B2), and Cyan for the attention attractor stimulus circle (RGBa: 0, 255, 255, 125) (Fig. B3). The interfaces were semi-transparent to establish a more natural integration of the virtual assets into the real world, which is represented by the alpha-value (a) that is lower than 255.

For *Planes on vehicle*, the alpha value was set to 180 (fig. B2) and the interface was locked to the vehicle's position in Unity. The plane was tilted (54.45°) to be positioned parallel to the vehicle's windshield (see fig. 4), and was move forward to hover above the front of the car so the bottom of the plane lined up with the front bumper, and the top of the plane ended 1.30 m above the dashboard.

The traffic lights consisted of a black box with dimensions 0.20 m wide, 0.38 m tall, 0.18 m deep. In the case of the *Pedestrian lights HUD*, the interface hovered at a distance of 1 m from the participant while being located 0.30 cm upwards and to the right. The interface had to be rotated by -148.78° around the Y-axis to face the participant and compensate for the off-centre placement. The *Fixed pedestrian lights* interface consisted of the same traffic light box placed on a pole with a height of 2.04 m. The traffic light was located across the road from where the pedestrian was positioned at 13 m.

The virtual fence consisted of two vertical walls parallel to the walking direction of the participants. A gate that opened in the yielding scenario and remained closed in the non-yielding scenario, faced the participants. The walls were 3.03 m tall, the gate and walkway was 2.50 m wide and the walls parallel to the walking direction had a length of 7.55 m. The zebra had the same length, and was a little bit narrower so the sides of the zebra were separated from the walls.

Figure B4 shows the assets hierarchy of Unity. As explained in section 2.3.2 many virtual objects were mapped to a new *AR Origin*. This *AR Origin* was positioned to correspond to the real world, allowing the researchers to anchor stationary items to it, and easily scale and move the AR assets in the real world.

- World-locked interfaces are contained in the HMI Locations folder, which is a subfolder of the *AR Origin* and therefore stationary to this origin and thus the real world.
- The vehicle-locked interfaces are mapped to the *sedan-car-01* asset, locking it to the vehicle's position.
- The head-mapped interfaces are mapped to *XRRig*, which contains the position and rotation of the Varjo XR-3 HMD, and therefore follows the participants' head movements and appears as a HUD.
- The *GazeTimers* folder and *TextLocation* item include the positions for the attention attractor stimulus and the post-trial question respectively.
- *Planes* consists of the rigid ground planes on which the virtual vehicle drives. These are not visible to the participant during the experiment.

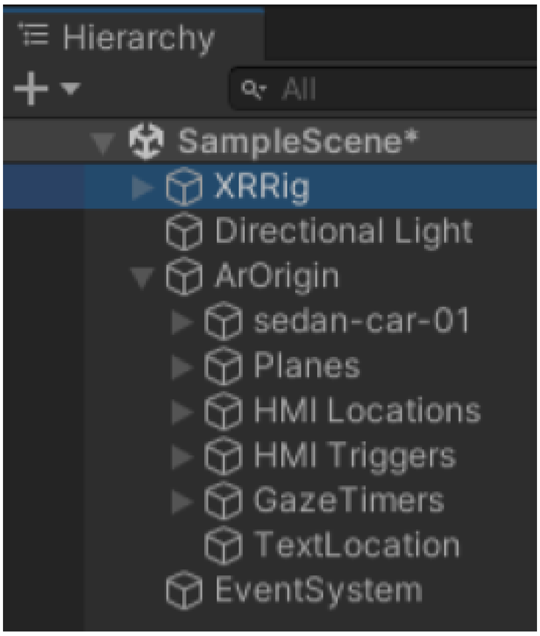

Figure C: AR assets mapped to AR Origin

- *HMI Triggers* holds the triggers that influence the vehicle's behaviour, such as *Interface activation, deceleration point* and *stopping point*.

D. Unity controls

Unity was set up in such a way that the experimenter could control all conditions from within Unity. The check boxes on the right under *Behaviour Script (script)* of the UI were used to select whether the vehicle would yield or not, which interface would be active, and where the attention attractor stimulus would be located. Moving one menu up under *Logger (script)*, the open text boxes provide the ability to insert the participant number and trial number, which in this case is participant 30 and trial 30.

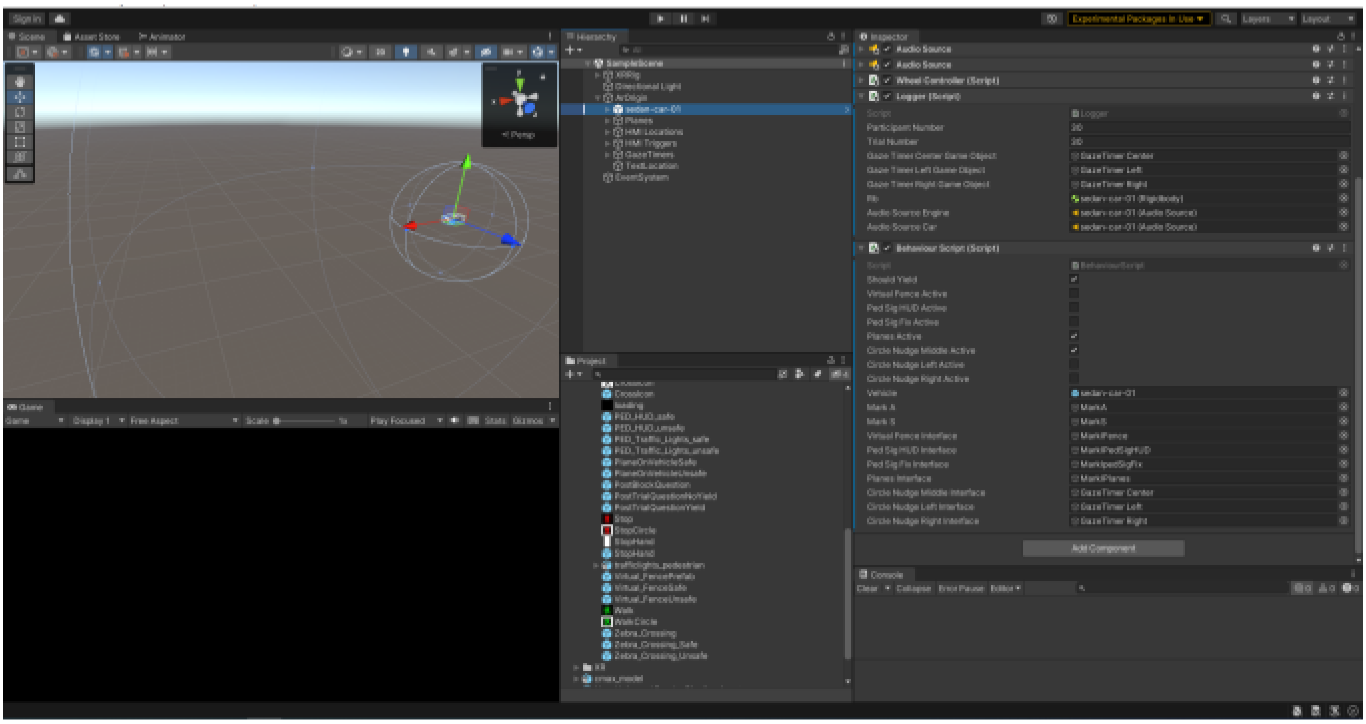

Figure D: Unity interface which experimenter used to control the experiment conditions.

E. Data processing

This step-by-step process described the method used to clean and sort the objective data recorded through the remote button.

1. Convert .csv log files into Excel files with separate columns (text-to-columns)
2. Cut off the rows before car velocity is greater than 30 km/h, which marks the moment the vehicle starts driving. Next move 501 rows down, which cuts off all data exactly 10.00 s after the start. This corresponds to the moment the intuitiveness question appears, marking the end of the trial.
3. Next, check the velocity of the vehicle in the last row that now exists:
    a. IF velocity = 0, move to *yield* folder
    b. IF velocity > 0, move to *non-yield* folder, because the vehicle continues driving.
4. From all the files currently in the *yield* folder, check if there are trials where the button has not been pressed at all. These will be deleted, as participants may have

forgotten to press the button at all, or the remote button had no connection to the receiver.

5. From all the files currently in the *yield* folder, check if there are trials where the button has not been pressed at all. These will be deleted, as participants may have forgotten to press the button at all, or the remote button had no connection to the receiver.
6. From step 3, copy log files into new folders to sort by block:
    a. BL / VF / PHud / PFix / PoV
7. Since remote button registers button press at <50 Hz, not all button presses are recorded at each discrete step. In the log files it seems that the refresh rate of the remote was approx 16 Hz. Thus, the solution is to take the max value per half metre. A half metre at 30km/h corresponds to 3/50th of a second, which corresponds to the remote's refresh rate of 16 Hz.
8. **Willingness to cross**
    a. The amount of ButtonPress = True values are summed across all log files at each PosX value resulting in Buttonpress_counts. This number is then divided by the total number of trials for each interface condition, resulting in the percentage of ButtonPress = True at each PosX value: PercentageButtonpress.
    b. Next, the maximum value of PercentageButtonpress is determined for each 0.5 m of distance, to accommodate for the slower refresh rate of the remote button. This is now plotted in a graph that also indicates interface activation, deceleration point and stopping point locations.
9. **Critical Gap**
    a. Search log files for first switch from TRUE to FALSE for ButtonPress boolean. Next index that row and find corresponding PosX value. That value is then plotted in the box plot.
10. **Decision Certainty**
    a. Search log files for first switch from TRUE to FALSE or vice versa for ButtonPress boolean, count and create box plot.

F. Experiment procedure

Experiment instructions

Introduction

We will conduct an experiment with an augmented reality headset. The headset uses cameras to capture the environment around you, and can project virtual objects onto the surrounding environment.

In this experiment you will be situated at the side of a road, and you will remain standing in one spot. You are a pedestrian and your goal is to cross the road. **<u>KEEP IN MIND</u>**: You will **NOT** physically cross the road, but indicate this by pressing a remote button, so you will not move or walk.

What to do

- Your main objective is to indicate whether you are willing to cross or not.
- You will do this by **pressing and holding** a button on a remote control when you feel safe to cross and releasing the button when you do not feel safe to cross.
- When the trial starts, you will see a circle in front of you, or to your left or right. You will now press the remote button briefly as a check.
- You will look into this circle for 1 second. The simulation will now start.
- A virtual vehicle will approach from the right.
- The vehicle will or will not stop for you to cross and will or will not indicate this via several communication interfaces.
- During the trial you will: **press and hold** the button on the remote control when you feel safe to cross and release the button when you do not feel safe to cross
- At the end of the trial, you will see a question appear in front of you. You will verbally answer the question to the researcher who will note your answer.
- You will encounter a total of 4 AR interfaces or no AR interface (baseline). After each block of 6 trials, I will ask additional open questions about the interface.

Your participation in the experiment is completely voluntary, and you are free to take off the headset at any time.

After the experiment you will complete the post-experiment questionnaire.

G. Statistical analysis

Interface	**Mean (*SD*) (yielding)**	**N**	**Mean (*SD*) (non-yielding)**	**N**
Baseline	47.27 (*19.85*)	23	28.21 (*23.60*)	28
Virtual Fence	56.00 (*22.82*)	23	25.76 (*22.93*)	28
Planes on vehicle	56.50 (*29.16*)	23	27.34 (*22.73*)	28
Ped. lights HUD	54.79 (*26.29*)	23	23.08 (*23.22*)	28
Fixed Ped. lights	54.84 (*25.77*)	23	23.83 (*18.74*)	28

Table G1: Willingness to cross percentage of trials distance averaged over all participants.

Interface	**Mean difference vs. Baseline (yielding)**	**p**	**Mean difference vs. Baseline (non-yielding)**	**p**
Virtual Fence	-8.715	.702	2.453	1.000
Planes on vehicle	-9.226	.854	0.872	1.000

Ped. lights HUD	-7.524	1.000	5.133	1.000
Fixed Ped. lights	-7.569	1.000	4.384	1.000

Table G2: Pairwise comparisons of the effect of the interface condition on the willingness to cross compared to baseline. The comparison number is determined by subtracting the willingness to cross percentage of the interface condition from the baseline condition. Therefore, a negative number indicates a higher willingness to cross percentage of the interface than the baseline condition.

Interface	Correlation (*sig.*) DC-Int. (Yielding)	N	Correlation (*sig.*) DC-Int. (Non-Yielding)	N
Baseline	**.405 (*.045*)**	25	-.287 (*.138*)	28
Virtual Fence	.085 (*.688*)	25	.044 (.825)	28
Planes on vehicle	-.002 (*.992*)	25	.281 (*.147*)	28
Ped. lights HUD	.297 (*.150*)	25	.018 (*.927*)	28
Fixed Ped. lights	-.091 (*.666*)	25	-.089 (*.653*)	28

Table G3: Correlation analysis between Decision Certainty (DC) and Intuitiveness (Int)

H. Post-block open questions

Baseline	Virtual fence	Fixed pedestrian lights	Pedestrian lights HUD	Planes on vehicle
Much more difficult than with interfaces. More doubt and uncertainty about my decision	Walls make you feel safe, as if they will protect you. Red state made me think twice.	I have to look back and forth so my final decision comes from the vehicle and not the traffic light.	Would be nicer if the difference between red and green was more obvious.	Interface on the car so it felt like a 1-to-1 interaction with the car.
A bit more difficult to decide because I miss the confirmation of the interface	Thought that red walls indicated for the vehicle to stop and zebra was an indication to cross for me.	Decision fully based on the vehicle. I used the interface to confirm my own decision.	I don't use the traffic light but look at the car to see if I can make it before the car arrives. Especially in this low-traffic scenario.	High trust in interface because the vehicle itself communicates the interface
Decision based on speed and deceleration	When looking at car, you do see it because it is large and visible	Did not see it the first time. Normally the traffic light would already be there before it activates	Interface placed far off to the side.	Did not use the interface to make a decision. Decision fully based on vehicle behaviour.
Absence of driver made it more difficult.	I still check if the vehicle slows down	Not easy to see because it is far away	Blocks a large part of FOV	A bit unusual. Had to get used to it.
Audio helps in the decision-making process	The zebra invited me to walk, even with the red walls.	If it’s green and I hear the car slowing down, then I am confident to cross	First I had to get accustomed to it. Once you know how it works, it is very easy	Nice that I only have to look in one direction
No driver in the car created more doubt. Normally I would make eye contact	The interface sort of blocks my view of the surrounding situation	Normally the traffic light will already be there but now it suddenly appears	With green it is clear that I can go. With red, I still try to go before the car arrives	Since it is connected to the car, it seems that it indicates what the car will do. For green I thought the car would drive.
I do wait a bit longer because there is no confirmation of the interface.	Unclear, because I thought the walls were indicators for the vehicle. Suggest to make the zebra red/green	I already made my decision based on the vehicle before the interface appears	Because it is a simulation, I fully trust that the vehicle’s behaviour will correspond to the interface	Icons are clear but decision mostly based on colour

Table H: Summary of the interface specific answers to post-block open questions