Towards Scalable eHMIs: Designing for AV-VRU Communication Beyond One Pedestrian

Debargha Dey
d.dey@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Arjen van Vastenhoven
a.v.vastenhoven@student.tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Raymond Cuijpers
R.H.Cuijpers@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Marieke Martens
m.h.martens@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Bastian Pfleging
b.pfleging@tue.nl
Eindhoven University of Technology
Eindhoven, The Netherlands

Figure 1: We conducted a VR-based study to test the scalability of several eHMI designs. How good are eHMIs at clear and unambiguous interactions in situations with multiple pedestrians? When an automated vehicle approaches more than one non-co-located pedestrians, can eHMIs effectively communicate whom the AV intends to yield to? If the eHMI communicates that the AV is yielding, is it clear to the participant whether the AV is yielding to them, or to the other pedestrian who is standing further down the road?

ABSTRACT
Current research on external Human-Machine Interfaces (eHMIs) in facilitating interactions between automated vehicles (AVs) and pedestrians have largely focused on one-to-one encounters. In order for eHMIs to be viable in reality, they need to be scalable, i.e., facilitate interaction with more than one pedestrian with clarity and unambiguity. We conducted a virtual-reality-based empirical study to evaluate four eHMI designs with two pedestrians. Results show that even in this minimum criteria of scalability, traditional eHMI designs struggle to communicate effectively whom the AV intends to yield to. Road-projection-based eHMIs show promise in clarifying the specific yielding intention of an AV, although it may still not be an ideal solution. The findings point towards the need to consider the element of scalability early in the design process, and potentially the need to reconsider the current paradigm of eHMI design.

CCS CONCEPTS
• Human-centered computing → Interaction design; Interaction techniques.

KEYWORDS
Automated vehicles, eHMIs, scalability, vulnerable road users, vehicle-pedestrian interaction

ACM Reference Format:
1 INTRODUCTION

With the advent of automated driving technology, the question of an effective interaction between automated vehicles (AVs) and other road users becomes critical in ensuring a successful and seamless integration into society. An external Human-Machine Interface (eHMI) is typically proposed to facilitate the interaction between AVs and vulnerable road users (VRUs) [1, 5, 11]. While research into AV-pedestrian interaction is rather new, a number of studies have been conducted on eHMIs, promising increasing trust and perceived safety [14, 20], improving crossing response time [19], improving the user experience of AV-pedestrian interaction [9], and mitigating ambiguity [6].

However, most of the research on eHMIs has been done in a simple setting, involving one vehicle and one pedestrian. While such studies are critical in understanding the fundamental principles that affect pedestrian behavior in interacting with vehicles, they do not necessarily reflect real-world traffic situations. Rather than being one-to-one, real-world traffic involves multiple entities with different intentions and different context, interacting with many vehicles at the same time. Thus, there is a need to extend the body of research into investigating the scalability of eHMI concepts in one-to-many and many-to-many AV-pedestrian interaction situations.

Prior work in the theoretical foundation of eHMI design [4, 8, 33] identified scalability as an important component of a good eHMI. A scalable eHMI is defined as one that is able to communicate with multiple traffic stakeholders (in this case, pedestrians) with clarity and unambiguity [8]. In the absence of an attention to the element of scalability in the existing corpus of work in eHMI research, this paper is a first step in investigating the design requirements of a scalable eHMI. To this end, we conducted a virtual reality (VR)-based study with three different existing eHMI concepts against the baseline condition of no eHMI. To this end, we conducted a virtual reality (VR) based study with three different existing eHMI concepts against the baseline condition of no eHMI, with two pedestrians. While two pedestrians still do not reflect the complexity of many real-world scenarios, it introduces the minimum criteria of scalability (i.e. more than one pedestrian) in a controlled experiment, and highlights several challenges for current eHMI designs. Our results reveal that in traffic environments with multiple pedestrians, simply communicating the yielding intention of an AV may not be clear enough, and may even lead to misunderstandings stemming from a pedestrian interpreting and misinterpreting a yielding message meant for someone else. This experiment serves as a test for several existing eHMI concepts with regard to scalability, and as a first step, shows that many existing concepts cannot communicate effectively and unambiguously in multi-pedestrian situations. Pedestrians need precise, directed communication regarding the AV’s yielding intention to make effective road-crossing decisions. On-vehicle eHMIs perform poorly in terms of offering clarity regarding which pedestrian the vehicle intends to yield to by virtue of not being able to precisely indicate where or for whom the AV would stop. Projection-based eHMIs can mitigate this ambiguity to some extent, but is still likely not an ideal solution. This has implications in terms of the potential need to alter the current paradigm of eHMI designs.

Contribution statement: In this paper, we present an empirical study that evaluates the scalability of eHMIs (the effectiveness of eHMIs in the context of communicating with two pedestrians) where a yielding message from an eHMI can lead to misunderstandings due
to different locations of the pedestrians. We found that some common designs of eHMI fall short when it comes to interacting with clarity when multiple pedestrians are present. Existing eHMIs that communicate yielding intention without specifying the individual it is yielding to can cause confusion in multi-pedestrian interaction scenarios. Our insights critically recommend that from a safety point of view, in such multi-pedestrian scenarios, it is rather advisable to not have an eHMI at all, rather than one that simply communicates yielding intention without specifying whom it is yielding to. This emphasizes the need to consider scalability for the design and development of future eHMIs. Scalability-sensitive intention-communication based eHMIs should look at precisely addressing the communication to specific recipients instead of making broad announcements about its yielding intention.

2 BACKGROUND AND RELATED WORK

Previous research involving eHMIs typically evaluated their effect on AV-pedestrian interaction in relatively simplistic scenarios. The literature review by Colley et al. [3] captures the state of the art when it comes to a scalable design for eHMIs and exposes this research gap. Of the 29 empirical studies identified by them (13 Virtual Reality (VR) studies, 7 video-based studies, and 9 real-world studies), four studies involved interaction with multiple traffic stakeholders at the same time: The work by Lee et al. [24] involved two cars approaching one pedestrian in a VR setting, with the pedestrian having been tasked to cross the road after the first car had passed. The VR-based study by Mahadevan et al. [26] explored group behavior with several co-located pedestrians, all of whom had the same intention to cross the road in front of an approaching vehicle. In the real-world study by Merat et al. [28], multiple pedestrians interacted with an autonomous pod in a shared space scenario (parking lot). However, no research has yet looked into evaluating eHMIs in situations where multiple pedestrians have to interact with AVs when they have different right-of-way situations or different implications for their crossing decision based on the AV’s yielding intention.

This is distinct from situations when multiple pedestrians are co-located with the same intention, or for whom the same rules apply. Previous research on proxemics (the interaction of individuals with each other and the environment depending on the physical distance or closeness between them) revealed that when pedestrians belong to a group, they tend to act with unity and do not make independent decisions [27]. A message from an AV to a number of pedestrians of the same group can react to the same message and do not necessarily need to be addressed individually – the same message applies to everyone, and the group behaves organically as a whole, or when one member of the group acts, the others follow. However, when multiple pedestrians are present in the environment who do not belong to a group (i.e. who make their decisions individually irrespective of the action of the others), the clarity of a message from an eHMI becomes important in helping them take effective decisions. In these situations, the scalability of an eHMI matters.

The difficulty of addressing each individual pedestrian increases with the complexity of the traffic conditions (and the number of pedestrians, especially in non-group settings). In such scenarios
(a) Potential misunderstanding of an eHMI: When the eHMI of the AV (A) communicates that it will yield to a pedestrian (Y) who is waiting to cross at a crosswalk, another pedestrian (X) who does not have right of way may intercept the message and misconstrue that the message is meant for them.

(b) Visualisation of an eHMI communicating yielding intent (top) and an eHMI communicating yielding intent + contextual information (bottom).

Figure 2: The scalability issue of eHMIs, and a potential eHMI solution

with multiple pedestrians having different intentions in traffic, it is critical that an eHMI is unambiguous to each traffic participant. Otherwise, eHMIs can lead to misinterpretation and a false sense of security [22]. An example of a potential problem arising from an unintended road user intercepting the communication from an eHMI and taking an action that can lead to a dangerous situation is shown in Figure 2a. In this case, the AV is obligated to yield to pedestrian Y who is waiting to cross the road at a crosswalk, and the eHMI communicates that the AV is yielding. But without additional context, pedestrian X, who does not have a right of way and who waits between pedestrian Y and the car, may intercept this message and combined with the slowing behavior of the vehicle mistakenly think that the message is meant for them. Rasouli et al. [33] highlight another example that if an AV slows down before turning at an intersection, a nearby pedestrian could incorrectly assume the vehicle is yielding to them. In a dynamic urban traffic environment, there could be many more situations where similar misinterpretations might happen.

This calls for a need for eHMIs that can communicate with enough clarity regarding the AV’s intention so that the pedestrians can interpret the safety of the situation unambiguously. This can be done by either individually addressing pedestrians with directed messages, or by providing further context about the yielding intention of the vehicle such as when, where, or for whom the AV will yield. For a more focused information regarding where the vehicle stops, Dietrich et al. [13] proposed road-projection eHMIs as a solution, which was validated by subsequent studies as an acceptable solution [25, 30]. Alternatively, a recent work by Dey et al. [9] proposed distance-dependent, contextual eHMIs that announce an AV’s yielding intention in general at a distance, and provides additional contextual information about the yielding intention when the AV gets closer to the pedestrian, which helps the pedestrian gain more insight regarding where, when, or for whom the AV intends to stop. Figure 2b shows this schematic concept. It was posited that such eHMIs improve the user-experience and provide more detailed and rich information to pedestrian [9], and might be more useful in terms of scalability than eHMIs that do not provide additional contextual information.

To this end, the goal of our concept and study presented in this paper are to investigate which eHMI design is suited to clearly communicate with multiple pedestrians in a non-group setting, and validate if providing additional contextual information improves the scalability of an eHMI compared to an ordinary eHMI without any additional information.

3 eHMI CONCEPTS

We based our eHMI concepts on promising existing designs from prior work. For this, we used adaptations of the distance-dependent eHMIs proposed by Dey et al. [9] and the road-projection-based eHMI proposed by Löcken et al. [25]. We evaluated four eHMI designs for their effectiveness in scalable communication, which are described below.

3.1 Bumper eHMI – Moving Light Bar on Bumper

As an iteration of prior ‘light band’ concepts [9, 10, 17], this eHMI is a one-dimensional light bar mounted on the bumper of the vehicle. While the AV cruises in automated mode, the eHMI glows in a solid, turquoise color. To indicate a yielding intention, two animated light segments repeatedly wipe sideways from the outside, flowing...
AutomotiveUI ’21, September 9–14, 2021, Leeds, United Kingdom
Debargha Dey, Arjen van Vastenhoven, Raymond Cuijpers, Marieke Martens, and Bastian Pfleging

Figure 4: Bumper+SA eHMI. The light bar on the bumper animates continually as the vehicle is yielding. The windshield shows the AV’s situational awareness by reflecting the position or the location of the pedestrian with respect to the car, and informs whom the car intends to yield to. As the car approaches, the visualization on the windshield moves accordingly to map to the pedestrian’s relative location. Images show the windshield display indicating situational awareness recognizing the presence of a pedestrian at – Left: the start of yielding. Right: the end of yielding.

Figure 5: Bumper+PB eHMI. The light bar on the bumper animates continually as the vehicle is yielding. Additionally, the windshield displays a progress bar that fills up in the windshield as an indication of when the vehicle will come to a complete stop. The linear growth of the windshield visualization – from empty to full – reflects the slowing of the car, and offers observers an estimate of when the car will come to a complete stop. Images show the windshield display indicating the progress bar at – Left: the start of yielding. Right: the end of yielding.

towards the center of the light bar (Figure 3). To indicate that the car wants to start driving again, the light bar returns to a steady glowing state. Although the sideways wiping eHMI conveys the vehicle’s yielding intention in general, it does not provide any additional information regarding whom the vehicle is yielding to or when/where it will come to a stop. In essence, the pulsating eHMI tells the pedestrian, “I am yielding”. The subsequent three eHMIs (Bumper + SA, Bumper + PB, and Bumper + SP) add the element of distance-based contextual information, either on the windshield, or on the road via a projection by indicating in one of three ways: for whom, when, or where the vehicle is going to stop. In all eHMI concepts, the base Bumper eHMI is always present.

3.2 Bumper+SA eHMI – Light Bar & Windshield “Situational Awareness” Display
This concept extends the light bar Bumper eHMI. In addition to showing the general yielding intention of the vehicle, a subsequent windshield eHMI activates when the car is at a certain closer distance from the pedestrian. This additional display provides directed information to the pedestrian(s) by informing them about the vehicle’s situational awareness regarding whom the vehicle is yielding to. In essence, this concept tells the pedestrians “I have seen you and I am yielding”. This concept is a replication of the Bumper+SA eHMI proposed by Dey et al. [9], which was inspired from other ‘tracker’ eHMIs [10, 32, 34]. The visualization on the windshield is a light segment: a vertical bar on the windshield lights up corresponding to the relative direction of the pedestrian with respect to the car to show that the car will yield to the recognized pedestrian (Figure 4). During this time, the light bar on the bumper continues to animate. As the car approaches the pedestrian, the light on the windshield moves across the windshield to reflect the change in the relative angle between the car and the pedestrian. When the car intends to start driving again, the windshield interface deactivates, and the light bar on the bumper returns to a solid glow. Thus, at a closer proximity, the windshield interface gives the pedestrian(s) more detailed information on the situational awareness, i.e., the fact that the vehicle is aware of the pedestrian, and the yielding intention of the car. This nuanced, directed information may help in a situation where there are multiple pedestrians by helping an individual understand specifically whom the AV intends to yield to.

3.3 Bumper+PB eHMI – Light Bar & Windshield “Progress Bar” Display
This concept is also an extension of the light bar Bumper eHMI. In addition to the light bar (to indicate yielding), additional information is provided on the windshield similar to Bumper+SA when the car is at a closer distance from the pedestrian. A progress bar gives the observers a time estimate of when the car will come
I am yielding and here is an estimate of when I will come to a complete stop. In essence, this concept tells the pedestrians “I am yielding and here is an estimate of when I will come to a full stop”. This concept replicates the Bumper+PB eHMI proposed by Dey et al. [9]: The windshield display shows a vertical band of light that originates from the center of the windshield and expands horizontally in a linear way as the vehicle approaches the stopping point (Figure 5). The state of the entire windshield being lit up is mapped to the car being at rest. The windshield interface stays lit (glowing steadily) as long as the car is yielding and the light bar on the bumper continues to animate. The windshield content diminishes in size and deactivates when the car is ready to start driving again, and subsequently, the light bar on the bumper returns to a solid glow. In a traffic situation involving multiple pedestrians, this additional information may help the pedestrian determine whom the AV intends to yield to by extrapolating the time information from the windshield display to an actual position on the road.

### 3.4 Bumper+SP eHMI – Light Bar & Street Projection Display

The final eHMI concept used in this study extends the Bumper eHMI with a road-projection. As with the previous concepts, the light bar activates to show the yielding intention of the vehicle. Additionally, the car indicates with a projected visualization on the road where the car intends to come to a stop. In essence, this concept tells the pedestrians “I am yielding and here is the place where I will stop”. This concept is an extension of the distance-dependent eHMI concepts proposed by Dey et al. [9], and is inspired by the F015 concept from Mercedes-Benz [2], and work by Lücken et al. and Dietrich et al. [13, 25]. The road projection eHMI informs the observer where the vehicle is going to come to a full stop by explicitly indicating the exact physical location (Figure 6). As AV starts to yield, the eHMI visualization projects two parallel lines that move away from the vehicle towards the exact location. Subsequently, the eHMI projects arrows that animate into a line which indicate the exact stopping location of the vehicle. The size of the projection diminishes as the vehicle approaches closer to the stopping point. Once the vehicle stops, the projection switches to a zebra crossing, and stays lit while the vehicle waits in front of the pedestrian, until just before the vehicle is ready to start driving again, when it deactivates. Showing the exact stopping point of a vehicle may help in a traffic situation involving multiple pedestrians by clarifying exactly where, and by extension – for whom – the vehicle intends to stop.

### 3.5 Research Question & Hypotheses

We test the presented eHMI concepts for their effectiveness in scalable communication when another pedestrian is present besides the ego pedestrian, and answer the following research question:

**Does adding specific information clarifying the context of a yielding message (for whom/when/where, etc.) enhance the clarity and scalability of communication of an eHMI compared to when no contextual information is provided?**

According to Habobovic et al. [16], a multi-pedestrian scenario such as the one represented in Figure 2a can lead to a faulty diagnosis of the traffic situation, and an eventual mistake in decision making due to incorrect analogy. This arises from the pedestrian thinking that a vehicle is slowing down for them, when it is actually slowing down for someone else. Consequently, we posit that adding contextual information to the eHMI about its yielding intention will clarify the background of its message with the following hypotheses:

**H1** When the AV yields to the participant, the eHMIs that provide contextual information – Bumper+SA, Bumper+PB, and Bumper+SP will reduce ambiguity and lead to a higher willingness to cross for the participant than the Bumper eHMI which does not provide additional contextual information.

**H2** When the AV yields to the pedestrian other than the participant, the eHMIs that provide contextual information – Bumper+SA, Bumper+PB, and Bumper+SP will reduce ambiguity and lead to a lower willingness to cross for the participant than the Bumper eHMI which does not provide additional contextual information.

Although each of the three contextual eHMIs added context by answering one of three questions – for whom, when, or where the vehicle will come to a stop – we did not have any rationale-driven hypothesis on which contextual eHMI might perform better. Therefore, we also considered an open-ended, exploratory question of which contextual eHMI was the best performer in terms of affecting the participant’s willingness to cross.

### 4 EXPERIMENT

We evaluated the proposed eHMI concepts in a VR-based within-subject experiment across five eHMI conditions: a baseline of No eHMI, and the four eHMI concepts described in Section 3. The experiment was approved by the ethical review board of the researchers’ institution(s).

#### 4.1 Task

In this VR-based experiment, the participants had to experience several trials of an AV approaching them while they assumed the role of a pedestrian intending to cross the road. While experiencing...
Within each block, the participant experienced three distinct vehicle behaviors: a) the vehicle yields to the participant, b) the vehicle yields to the other pedestrian, and c) the vehicles does not yield at all. For each eHMI condition and each vehicle behavior, the participant experienced five trials. Thus, each block consisted of 15 trials (3 vehicle behaviors × 5 trials per behavior). With the 5 blocks of eHMI concepts, each participant experienced a total of (5 × 15 = ) 75 trials. The occurrence of the trials within a block was also presented in a counterbalanced order. The study design is shown in Table 1.

For each trial, the vehicle followed the same path (see Figure 7), appearing from a bend in the road 100 m away and approaching the pedestrian at a speed of 50 km/h. The vehicles started at the left side of the pedestrian, at a distance of more than 100 m away. We added a non-yielding behavior to avoid the learning effect that the vehicle always yields to a pedestrian. In the non-yielding conditions, the vehicle drove at a constant speed of 50 km/h while being visible. In all yielding conditions (either to the participant or the other pedestrian), the car started to slow down at 43 m from the pedestrian, and came to stop 3 m from the pedestrian, leading to a total deceleration distance of 40 m. For both yielding conditions, the vehicle decelerated in around 5.3 seconds, following a literature-supported normal braking deceleration rate of 2.4 m/s² [7]. The vehicle waited for three seconds in front of the pedestrian, and then drove away.

The distance between the participant and the other (virtual) pedestrian was 10 m. This distance was chosen as an optimal distance to be not so close as to stimulate group behavior, but close enough that a yielding message from an eHMI of the approaching AV could mean either of the two pedestrians. This distance was further verified as an appropriate separation in a pilot study by the experimenter.

Due to the COVID-19 related restrictions, the experiment could not be performed in the VR laboratory of the university, but it was deployed as a remote VR study where participants used a mobile VR app developed in Unity which was running on their own mobile phone embedded into a Google Cardboard headset2. In the VR environment, the participant was able to look around freely in a 360°. They were able to indicate their willingness to cross (binary: pressed - yes, released - no) using the button on the top-right corner of the Google Cardboard headset.

### 4.3 Procedure

Participants received a Google cardboard headset ahead of the experiment, as well as an Android application with a virtual environment which they installed on their mobile phone prior to the experiment. The participants were asked to look around in their virtual environment and understand that they will be interacting with the AV in the presence of another pedestrian. They were also demonstrated the use of the Google Cardboard button as the input method for their willingness to cross – they were asked to keep the button pressed as long as they felt safe to cross the road, and release it when they felt unsafe. All participants tried out a demo before the experiment in which they got to know the environment.

---

1. https://unity.com/, last access 2021-05-10

---

Table 1: Study design: All participants first experienced the ‘No-eHMI’ block. The remaining four eHMI blocks followed in a counterbalanced order. We also counterbalanced the order of stimuli within a block.

<table>
<thead>
<tr>
<th>Trial #</th>
<th>eHMI Concept</th>
<th>Vehicle Behavior</th>
<th># Repetitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 15</td>
<td>No eHMI (Block 1)</td>
<td>Yielding to participant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yielding to other pedestrian</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Yielding (50km/h constant)</td>
<td>5</td>
</tr>
<tr>
<td>16 – 30</td>
<td>Bumper (Block 2)</td>
<td>Yielding to participant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yielding to other pedestrian</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Yielding (50km/h constant)</td>
<td>5</td>
</tr>
<tr>
<td>31 – 45</td>
<td>Bumper+SA (Block 3)</td>
<td>Yielding to participant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yielding to other pedestrian</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Yielding (50km/h constant)</td>
<td>5</td>
</tr>
<tr>
<td>46 – 60</td>
<td>Bumper+PB (Block 4)</td>
<td>Yielding to participant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yielding to other pedestrian</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Yielding (50km/h constant)</td>
<td>5</td>
</tr>
<tr>
<td>61 – 75</td>
<td>Bumper+SP (Block 5)</td>
<td>Yielding to participant</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yielding to other pedestrian</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Yielding (50km/h constant)</td>
<td>5</td>
</tr>
</tbody>
</table>

---

Figure 7: Setup of the experiment: The movement path of the AV is shown with a dotted yellow line, starting from the left side. The participant approached the road at position 1, while the second pedestrian was located at position 2. The white arrows indicate the braking paths for the two different yielding conditions.
learned how to use the button on the Google Cardboard to indicate their willingness to cross, and got familiar with instructions and the setup in VR.

The entire experiment took approximately one hour in which the experimenter coordinated with the participant and guided them through the process via Microsoft Teams. After explaining the study setup, the experimenter ensured that the participant’s phone was set up properly, by verifying a working Internet connection, a sufficient battery level, and asking the participants to switch off their notifications. Subsequently, the experimenter provided the password for the Android app, so that participants could enter the experiment section of the app. This password protection ensured that participants did not use the App beforehand. After the participant put on the cardboard headset, the experiment began with the start of Block 1.

At the start of every block, the App explained the eHMI in that block and showed a preview of the AV with the eHMI in order to familiarize the participant with the eHMI. We wanted to ascertain that the pedestrians understood the eHMIs presented in the study, and the results of their responses were an accurate measure of the effectiveness of the eHMI and not its intuitiveness. After the explanation and the preview, the participant proceeded with the experimental trials in the block. At the end of every block, the participant took off their headset, and filled a questionnaire indicating their experience of the interaction, and their understanding of the behavior of the approaching vehicle. Between each block, the experimenter checked whether the participant was feeling well, experienced any signs of nausea or dizziness, and felt comfortable to proceed with the study. After the fifth (last) block and the corresponding questionnaire, the experiment concluded with a short semi-structured interview/discussion with the participant regarding their experience.

4.4 Measures

In this study, we used two broad measures to evaluate the different eHMI concepts. The first measure was the quantitative data of **Willingness to Cross** from the Google Cardboard button presses, which was used as an objective measure for the clarity of the eHMI’s message, as past research has shown that a pedestrian’s willingness to cross can be used as a surrogate measure for pedestrians’ feeling of safety or trust around an approaching vehicle [1, 36]. The binary willingness to cross data (pressed button indicating willingness to cross = 1, and released button indicating no willingness to cross = 0) were stored in an online PlayFab database. Each instance of a change in the participant’s willingness to cross was recorded with a timestamp and the corresponding distance of the front bumper of the vehicle from the location of the participant.

The second measure was the subjective perception of the eHMI concepts as revealed by the questionnaire responses. For this purpose, we used the user experience scales posited by Ackermans et al. [1], and consisted of three dimensions: (1) the **positive experience** of the participants with the vehicle, (2) the **general intent understanding** – indicating whether the participant understood the overarching yielding or non-yielding intentions of the vehicle, and (3) the **specific intent understanding** – indicating whether the participants understood when vehicle was yielding specifically to them, or for the other pedestrian. Each of these three dimensions were answered over questions which were measured on a seven-point Likert scale.

4.5 Participants

We recruited university students and staff with (corrected to) normal vision using the university experiment participation database (N = 36, 14 female, 22 male; M_age = 23.3, SD = 2.66). Each participant read and gave their informed consent for participating in the study, and was compensated for their time according to the policy of the university.

5 RESULTS

We analyzed the quantitative measure (Willingness to cross) and qualitative measures (positive experience, general, and specific intent understanding ) separately and present the results in the following sections.

5.1 Willingness To Cross

The Willingness to Cross data (whether a person was willing or unwilling to cross at any given moment with respect to the time-to-stop or the time-to-arrival of the car) was sampled every 0.25 seconds. For analysis, we focused on the time window spanning 2 seconds until the vehicle stopped or arrived at the pedestrian’s location. This is because this is arguably the most critical moment in making a road-crossing decision and any faulty diagnosis about the vehicle’s intention in this time window is highly meaningful.

5.1.1 Yielding to the participant. Figure 8a shows the average willingness to cross for vehicles yielding to the participant. In this condition, the ideal scenario is that participants notice early on that the vehicle is yielding to them, so that they feel safe earlier, cross the road early, and contribute to more efficient traffic. A high willingness to cross early before the vehicle stops is potentially an indicator of a good eHMI. The graphs shows a clear distinction between the baseline and the other four eHMI conditions. The baseline condition with No eHMI yielded the lowest willingness to cross. The Bumper+SP eHMI shows the highest overall willingness to cross scores, while the other three eHMI scores in the middle.

A logistic regression run over the last two seconds of stopping indicated significant effects for all eHMI conditions, in which all the eHMIs (Bumper, Bumper+SA, Bumper+PB, and Bumper+SP) all significantly increased the willingness to cross compared to the baseline condition (p < 0.001). The model had an R-squared of 0.2349, and the Hosmer-Lemeshow test rejected a lack of appropriative fit (p = 0.2993).

A pairwise comparison was run to compare the different conditions in more detail. The comparisons revealed that all eHMIs differ significantly (p < 0.001) from each other, except for the Bumper and the Bumper+SA eHMIs. These results are summarized in Table 2.

5.1.2 Yielding to the second pedestrian. Figure 8b plots the average willingness to cross for vehicles yielding to the other pedestrian. In this condition, the participants should notice early on that the
vehicle is not going to yielding to them, so they can quickly decide not to cross, reflected by a low willingness to cross. Subsequently, a lower willingness to cross before the vehicle arrives is an indicator of a good eHMI in clarifying that the yielding message is not meant for the participant. The graph shows that generally the different eHMI conditions cause a very similar behavior. In the two seconds before arrival, there seems to be some differences, which were subjected to statistical analysis.

The logistic regression indicated significant effects for all eHMI conditions when comparing them to the baseline condition. With respect to the No eHMI condition, three eHMI concepts led to a significantly higher willingness to cross: Bumper \((p < 0.001)\), Bumper+SA \((p < 0.001)\), and Bumper+PB \((p < 0.001)\). In contrast, Bumper+SP led to a significantly lower willingness to cross compared to the baseline No eHMI condition \((p = 0.010)\). The model had an R-squared of 0.2701, and pairwise comparisons were conducted for further investigation (Table 2).

Results of the pairwise comparisons support the findings from the logistic regression model that the Bumper, Bumper+SA, and Bumper+PB eHMIs all significantly increase the willingness to cross in the two seconds before the vehicle arrives at the participants’ position when the vehicle was not yielding to the participant, but instead was yielding to the other pedestrian. These eHMIs therefore perform worse than the No eHMI condition at this moment, as their communication slightly increases participants willingness to cross, whereas an effective eHMI should do the opposite. The baseline of No eHMI therefore outperforms these three eHMI conditions. The use of the street projection in the Bumper+SP eHMI however shows a significantly lower willingness to cross in this time window.

We do not report the findings of the non-yielding condition in detail, as this is not the focus of this study, and this behavior was added to prevent a learning effect regarding the vehicle’s behavior. However, the results for non-yielding condition are summarized in Figure 9 and Table 2 along with that of the two yielding behaviors.

5.2 Questionnaire data

We tested three dimensions subjectively in the questionnaire: (1) positive experience, (2) general intent understanding, and (3) specific intent understanding (see Figure 10). For each dimension across each of the 5 eHMI conditions, repeated-measures ANOVA was executed to test for the effect of the eHMI. All eHMI conditions had significant effects on positive experience \(F(4, 137) = 21.29, (p < 0.001)\), general intent understanding \(F(4,137) = 38.88, (p < 0.001)\), and specific intent understanding \(F(4,137) = 29.96, (p < 0.001)\).

We ran pairwise comparisons per scale to examine the differences between the eHMI conditions (see Table 3). Participants rated...
Table 2: Pairwise comparisons of mean willingness to cross of the last two seconds before arrival or stopping of AVs for all eHMI conditions, displayed per yielding behavior. Significant effects at a Bonferroni corrected confidence level of 0.005 (for 10 comparisons) are highlighted in bold and italics, and presented along with effect size Cohen’s d. BL = baseline (No eHMI), BU = Bumper, SA = Bumper + Situational Awareness, PB = Bumper + Progress Bar, SP = Bumper + Street Projection.

Table 3: Pairwise comparisons of mean questionnaire scale scores. Significant effects at a Bonferroni corrected confidence level of 0.005 (for 10 comparisons) are highlighted in bold and italics, and presented along with effect size Cohen’s d. BL = baseline (No eHMI), BU = Bumper, SA = Bumper + Situational Awareness, PB = Bumper + Progress Bar, SP = Bumper + Street Projection.

5.3 Evaluation of Hypotheses

Results from both the quantitative and qualitative analyses show that although there was a change in objective willingness to cross and perception of intent for eHMI conditions compared to the baseline No eHMI condition, no blanket statement can be applied to all contextual eHMIs as a whole.

We observed that when the AV was yielding to the participant, there was no significant difference between the non-contextual Bumper eHMI and the contextual Bumper+SA eHMI. However, both Bumper+PB and Bumper+SP eHMIs performed better than the non-contextual Bumper eHMI, and the effect of the street-projection (Bumper+SP) was very high. Thus, hypothesis H1 – which suggests that when the vehicle yields to the participant, the proposed contextual eHMIs will increase their willingness to cross compared to a non-contextual eHMI – cannot be fully supported.

For the cases when the AV stopped for the other pedestrian, we saw that the Bumper, Bumper+SA, and Bumper+PB eHMI conditions actually increased the participants’ willingness to cross compared to the No eHMI baseline condition. This is the opposite of an ideal outcome. Only with the projection-based Bumper+SP eHMI did the participants realize that the AV did not mean to yield to them, but for the other pedestrian, and respond with a lower willingness to cross. Thus, hypothesis H2 – which suggests that when the vehicle yields to the second pedestrian other than the participant, the evaluated contextual eHMIs will highlight this and decrease their willingness to cross compared to a non-contextual eHMI – cannot be fully supported.

6 DISCUSSION

Our results show that context-dependent eHMIs do not universally lead to a highly scalable eHMI design. The original assumption was that adding contextual information to a simple yielding intention
message in terms of for whom, when, or where an AV might be yielding may lead to a better eHMI design. We conjectured that providing additional contextual information would help an AV communicate more effectively, clearly, and unambiguously when multiple, non-co-located pedestrians are present. To this end, we used a non-contextual Bumper eHMI (which simply indicated the yielding intention of the vehicle), and compared it to other on-vehicle eHMI concepts that communicate contextual information. However, not all contextual information is the same, and the ways they are presented in an eHMI has implications for its clarity in a scalability context.

**Effectiveness of Street-projection based eHMI**

When the AV yielded to the participant, the Bumper + PB and Bumper + SP eHMIs increased the willingness to cross compared to the Bumper eHMI. The Bumper+SP eHMI performed particularly well. However, the Bumper+SA eHMI, despite providing additional contextual information, did not perform better than the non-contextual Bumper eHMI. Nevertheless, all eHMI conditions performed better than the No eHMI baseline condition, which is in line with previous findings that eHMIs clarifying an AV’s yielding intention tend to increase pedestrians’ willingness to cross [5, 23].

However, the outcomes are different when the AV yields to the other pedestrian waiting further down the road from the participant. The eHMI of the AV activates at a moment when the participant still sees the vehicle’s yielding intention, and at this point a road-crossing decision is contingent upon interpreting whether the yielding message is meant for the participant or the other pedestrian. In this situation, a good eHMI would clarify that the yielding message is not meant for the participant (first pedestrian), and this is done only by the Bumper+SP eHMI. The non-contextual Bumper eHMI actually increased the participants’ willingness to cross slightly compared to the No eHMI baseline condition, which indicated that the yielding message was erroneously interpreted as directed to the participant.

Surprisingly, the same outcome applied for the Bumper+SA and Bumper+PB contextual eHMIs. This shows that these two eHMIs were not able to clarify and disambiguate the fact that the AV was not actually yielding to the participants.

It appears that the street-projection based Bumper+SP eHMI performed better than the other contextual eHMIs in terms of scalability. Qualitative data from the questionnaire responses are also in line with these findings, and overall, participants rated their experience with the Bumper+SP eHMI as the most positive.

**The critical importance of focusing the yielding message of an eHMI to a recipient**

An interesting observation is that although the participants’ increase in willingness to cross was statistically significant for some of the eHMI conditions when the vehicle was yielding for the other pedestrian, the effect size was rather small. This leads us to infer that the driving behavior of the vehicle communicated the intent of the vehicle (that it was not yielding to the participant but the other pedestrian) more than the eHMI.

This further corroborates prior findings that the vehicle kinematics are a critical contributor to road-crossing decisions [12, 29]. This is also in line with previous findings that when the message from an eHMI is in disagreement with the vehicle’s driving behavior, pedestrians tend to fall back on the behavior to make their crossing decision [11].

That said, the effect of the eHMIs in increasing the willingness to cross should not be ignored. In these cases, the eHMI indicated the yielding intention of the vehicle, but failed to accurately specify that it was not meant for the participant, which led to an increase in their willingness to cross. Thus, if not designed carefully, pedestrians may still be prone to accept a non-specific yielding message from an eHMI as an invitation to cross, and this can lead to misunderstanding.

This is also in line with findings by Kaleefathullah et al. [22] that eHMIs can be misleading. An increase in willingness to cross at the wrong moment caused by an eHMI, even for a few individuals, may have catastrophic consequences in real world traffic, and undermines the initial goal of AVs to reduce traffic fatalities and injuries. Therefore, an eHMI causing people to misinterpret a yielding message that is not meant for them is a highly undesirable outcome, and care should be taken in the early phases of eHMI design to account for such behaviors.

Prior findings by Dietrich et al. [13] showed that when a message sent by an AV is not directed, confusion is likely to arise. Similar results were highlighted by Hensch et al. [18], where participants noticed light signals of AVs in a parking area, but misinterpreted whether it was directed to them. This study corroborates these findings in a multi-pedestrian setting, and once again points out the need for eHMIs to be clear an unambiguous in their communication.

The goal of scalability – simply the ability to communicate to multiple pedestrians at a time – is not enough on its own. A good eHMI must not only be scalable, but also be able to individually and unambiguously address whom a yielding message is relevant for. This dimension of an eHMI was identified as the *Communication*
Resolution of an eHMI in literature [8], and in this study, we provide an empirical proof of its importance.

Why did the other contextual eHMIs not work?
The rationale behind the contextual eHMI – that adding further information regarding for whom, when, or where the AV is stopping would increase clarity – is logically sound, and it was validated in prior work with one-on-one interaction [9]. The question arises then, why did the Bumper+SA (“to whom”) and Bumper+PB (“when”) eHMIs not perform well when the element of scalability (multiple pedestrians) was introduced? A common factor tying these two eHMIs is that both provided the contextual information with an on-vehicle windshield display, which made it difficult to map the contextual information provided to the real environment.

The situational awareness (SA) display provided a directed message with a ‘tracking’ light that reflected the relative position of the pedestrian the AV was yielding to with respect to the car. However, with multiple pedestrians present along the side of the road, there was room for ambiguity in the interpretation of the visualization of whom the AV was yielding to represented by the tracking light. As the complexity of the traffic situation increases with multiple co-located pedestrians, there will be ample room for misinterpretation whether the tracker on the situational awareness based display identified them or another other nearby pedestrian. This points to a design challenge in the numerous situational-awareness tracker based eHMIs that have been proposed in the past [10, 32, 34, 35].

Similarly, it is possible that the windshield display of a Progress Bar eHMI (Bumper+PB) was difficult for them to map the status of the progress bar with an actual stopping point on the road, as discussed in prior work [9]. In order to act as an effective “timer” to indicate when the AV would stop, a pedestrian would need to extrapolate the information about the extent of the progress bar on the windshield along with the current distance and speed of the vehicle in real time. Thus, a growing or shrinking visualization as a representation of an AV’s stopping moment is not an ideal mechanism to communicate nuanced information about an AV’s yielding behavior [10, 17].

Although the street projection based eHMI concept showed where the vehicle would stop, in highlighting a spot on the road, it was also able to answer whom the car was stopping for by extension. While the other two contextual concepts (Bumper+SA and Bumper+PB) attempted to communicate the additional information such as for whom and when the car will stop, they were not able to accomplish this with complete clarity. There was interpretation involved in understanding whether it was safe to cross the road for a certain pedestrian. This shows that although prior work highlighted the improved user experience of additional contextual information in eHMIs [9], simply providing this contextual information without disambiguating whom this information is meant for is not helpful in scalable communication.

Implications for eHMI Design and Future Work

One of the biggest implications of this study is that in situations with multiple pedestrians, it is critical than an eHMI – if used – should communicate clearly and unambiguously whom a yielding message is meant for in order to prevent the chance of miscommunication. Otherwise, from a safety perspective, an AV without any eHMI is more advisable than an AV with an eHMI that only communicates yielding intention without specifying the detail of its yielding intention. Although an eHMI specifying simply a yielding intention without further context can improve crossing behavior in one-on-one interactions, in more complex situations with multiple pedestrians, it can be confusing at best, and dangerous at worst.

The current design paradigm of visual eHMIs follow largely on-vehicle solutions. The interface lies on the body of the vehicle, and the communication occurs at the location of the eHMI. However, as this study showed, when the visualization of the AV’s yielding behavior is located on the body of the vehicle, it is difficult to communicate some aspects of contextual information with enough clarity and resolution. This can be seen as a future design challenge – we need to find a better way to communicate contextual information for eHMIs.

Street-projection-based eHMIs are able to take the location of the communication away from the car, and on to the road where the actual interaction takes place, which increases the clarity of the communication. Prior work also supports the effectiveness of...
projection-based eHMIs [25, 31]. However, there are also design challenges with projections, as the clarity of the projected visualizations are highly dependent on the environment (such as ambient light, weather conditions, and the nature, texture, and cleanliness of the road surface).

Alternatively, the specific communication regarding the yielding intention of an AV and what it means for individual pedestrians or road users may be taken away from the car entirely. With an increasing number of pedestrians, high resolution communication (individually and unambiguously identifying each pedestrian) becomes more challenging. Thus, for highly directed communication with a consideration for scalability, it is interesting to explore extra-vehicular solutions to facilitate AV-pedestrian interaction. Examples of such solutions could be smart infrastructure (where the communication takes place on the infrastructure such as smart roads or traffic lights), or personal devices (where the communication takes place on wearables or other smart devices). However, each solution has other associated drawbacks and challenges (such as associated cost and extent of deployment for effective use). Therefore, the correct course for scalable AV-pedestrian interaction must be investigated further.

The biggest takeaway of this work is that when an eHMI announces a vehicle’s yielding intention in multi-pedestrian environments, it is not sufficient to simply know that the vehicle is yielding, but particularly to whom it is yielding. Thus, minor modifications, iterations, or refinement to the status quo of the design patterns of eHMI concepts that simply announce the vehicle’s yielding intention will not be enough to accomplish unambiguous and scalable communication. The research community needs to look at novel solutions that – while communicating a yielding message – disambiguate whom the vehicle is yielding to. Alternatively, focus should be brought on designing and evaluating other forms or communication that does not explicitly communicate the yielding intention of the vehicle, but rather aspects such as the perception, belief state, or driving state of the vehicle (such as through highlighting vehicle kinematics).

7 CONCLUSION
This paper presents a VR-based experiment that explored AV-pedestrian interaction with two pedestrians to investigate the scalability of eHMIs beyond one-on-one interactions. Four types of eHMIs – one which simply communicated the AV’s yielding intention, and three others that added various further context about the yielding intention – were compared to each other and to a baseline condition of no eHMI. Our results show that for scalable communication with multiple pedestrians, clear, unambiguous communication is critically necessary to prevent miscommunication. Offering the context of an AV’s yielding intention through street projection yielded highly positive results. Other forms of contextual information provided via on-vehicle eHMIs were not as successful. The insights highlight the importance of considering the scalability aspect of AV-pedestrian communication early in the design process, and show the potential of extra-vehicular eHMI designs where the message is communicated close to the yielding location.

REFERENCES


[10] Debargha Dey, Marieke Martens, Chao Wang, Felix Ros, and Jacques Terken. 2018. Interface concepts for intent communication from autonomous vehicles to vulnerable road users. In Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. 82–86.


