Portobello: Extending Driving Simulation from the Lab to the Road

Fanjun Bu
Cornell Tech
New York, U.S.
fb266@cornell.edu

Stacey Li
Cornell Tech
New York, U.S.
sl3326@cornell.edu

David Goedicke
Cornell Tech
New York, U.S.
dg536@cornell.edu

Mark Colley
mark.colley@uni-ulm.de
Ulm University
Ulm, Germany
Cornell Tech
New York, U.S.

Gyanendra Sharma
Woven Planet
Los Altos, California, USA
gyanendra.sharma@woven-planet.global

Wendy Ju
Cornell Tech
New York, U.S.
wendyju@cornell.edu

Figure 1: Portobello enables *platform portability*, so that the same study can be run on in-lab (left) and on-road (right) driving simulation platforms.

ABSTRACT

In automotive user interface design, testing often starts with lab-based driving simulators and migrates toward on-road studies to mitigate risks. Mixed reality (XR) helps translate virtual study designs to the real road to increase ecological validity. However, researchers rarely run the same study in both in-lab and on-road simulators due to the challenges of replicating studies in both physical and virtual worlds. To provide a common infrastructure to port in-lab study designs on-road, we built a platform-portable infrastructure, Portobello, to enable us to run twinned physical-virtual studies. As a proof-of-concept, we extended the on-road simulator XR-OOM with Portobello. We ran a within-subjects, autonomous-vehicle crosswalk cooperation study (N=32) both in-lab and on-road to investigate study design portability and platform-driven influences on study outcomes. To our knowledge, this is the first system that enables the twinning of studies originally designed for in-lab simulators to be carried out in an on-road platform.

CCS CONCEPTS

• **Human-centered computing → Systems and tools for interaction design**; *Interaction design process and methods*; • **Computing methodologies → Simulation tools**; **Simulation environments**.

KEYWORDS

Human-Autonomous Vehicle Interaction, Driving Simulations

ACM Reference Format:
Fanjun Bu, Stacey Li, David Goedicke, Mark Colley, Gyanendra Sharma, and Wendy Ju. 2024. Portobello: Extending Driving Simulation from the Lab to the Road. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3613904.3642341

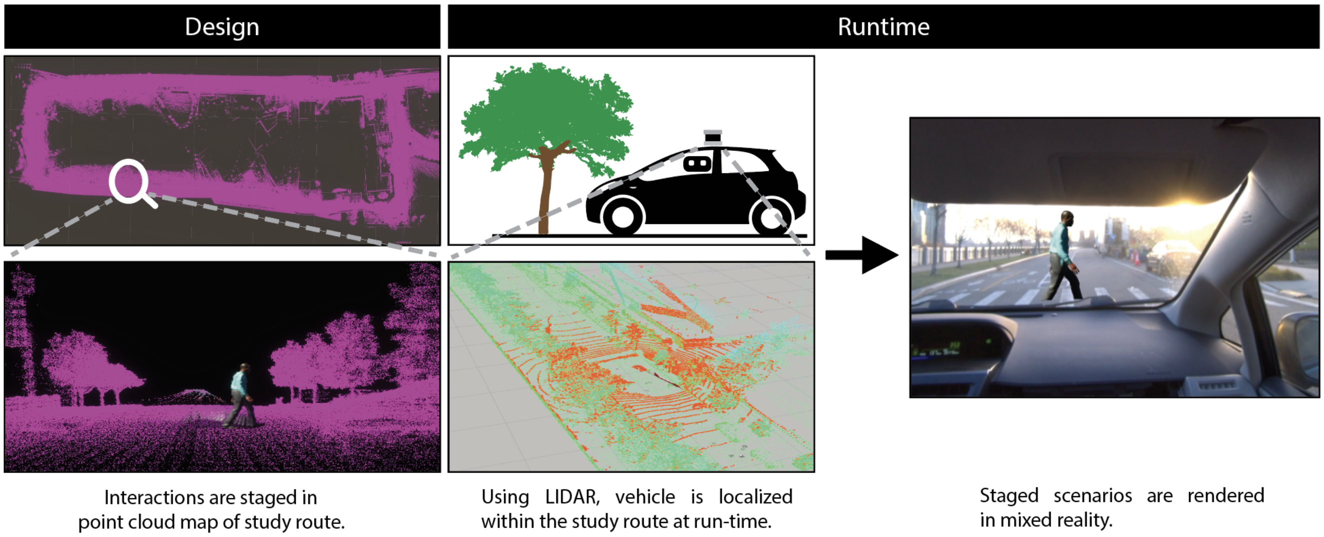

Figure 2: Complete execution pipeline using the Portobello infrastructure. During the design phase, researchers can drag and drop virtual objects on the point cloud map. At run time, the LiDAR-based navigation system locates the vehicle's position within the same map and sends the relative vehicle position to Unity. As a result, passengers wearing the video-see-through headset can see virtual objects at their corresponding real-world locations.

1 INTRODUCTION

Driving simulators play a critical role in human-centered automotive research applications because they allow people to experience different driving scenarios in a safe and repeatable fashion. Researchers have full control over the simulation setup and can program complex events in virtual environments. However, one of the major challenges for driving simulation has always been the difficulty of replicating the inertial forces and vehicle dynamics present in on-road driving [43]–even the most high-end motion platform driving simulators only replicate a fraction of the forces felt in real-world driving [19]. These forces are more critical in testing automated vehicle (AV) scenarios, where participants are often given non-driving-related tasks that keep them from looking at the screens of the simulations (e.g., [56]). Without the physical sensation and the visual engagement of the simulation environment, the immersiveness of the environment drops, making the evaluation of novel user experience and interaction techniques such as VEmotion [5] or SoundsRide [30] difficult. One way to address this problem is to incorporate driving simulation into a real vehicle driving on real streets (on-road simulators) [2].

On-road simulators are possible due to the maturity of XR, where digital displays blend reality and virtuality to increase the level of a user's immersion. The use of XR-in-the-car using simulated or actual dash-mounted heads-up displays have been explored by prior researchers such as Tonnis et al. [53], Kim and Dey [32], Schall Jr et al. [49], Ghiurău et al. [15], McGill et al. [39], Colley et al. [9], von Sawitzky et al. [54], Narzt et al. [41], and Bark et al. [3].

Despite this influx of on-road simulators under development, it remains challenging to migrate studies from in-lab simulators to on-road simulators. From a study design perspective, the key difference between in-lab and on-road simulators is that with in-lab simulators, designers have full access not only to the virtual vehicle but also to the virtual environment. The position and orientation of every brick are available to the designer with high precision, which makes event staging as simple as dragging-and-dropping modules into a map. However, researchers do not have easy access to objects outside the vehicle in on-road simulators. As a result, no on-road driving simulation system to date considers the surrounding context outside the vehicle for event staging, which limits the range of applicable studies. To replicate in-lab simulation on on-road platforms, access to the out-of-vehicle surrounding context is crucial.

In this paper, we describe a novel driving simulation infrastructure called Portobello, which enables *platform portability* in virtual driving simulation by incorporating localization technology and software from robotics. We define *platform portability* as the ability to run the same study on different (in-lab vs. on-road) platforms, an approach which we refer to as the twinning of studies. For this demonstration, we extended XR-OOM, a state-of-art XR driving simulation [16], to support on-road, location-based event staging (see Figure 2.) To validate Portobello's *platform portability*, we developed a within-subjects crosswalk-cooperation study (N=32) to be run on both an in-lab fixed-based vehicle chassis driving simulator and the on-road driving simulator built on top of the Portobello system [55]. As part of this work, we investigated how the different simulation platforms may affect the design process and results of user studies. The primary contribution of this work is the technical infrastructure system of Portobello, as validated by the proof-of-concept study. In addition, we provide a definition of platform-portability in virtual reality driving simulation and contribute insights into the process needed to develop twinned studies whose deployment is intended across multiple platforms. Finally, we demonstrate the relative strengths of different study platforms in the course of automotive research.

2 RELATED WORK

Driving simulation platforms are intended to be proxy environments that enable researchers to conduct studies where real-world experiments are dangerous or impossible. The standard for such platforms is face validity [14]: when participants take a simulation seriously, researchers can have greater faith that the study's results will be applicable to the real world. It is more important that the simulation allows participants to behave as if they are in a realistic setting than it is for the simulation itself to replicate reality in fine detail. In-lab and on-road simulators provide different approximations of driving scenarios.

2.1 In-Lab Driving Simulators

In-lab driving simulators are used to test interactions between drivers and the vehicle, the driving environment, and other in-world agents [27]; simulation allows researchers to observe in-vehicle behaviors safely, without physical risk to study participants [52]. Simulations can be implemented with low fidelity but still provide insights into how drivers will react sitting in a real vehicle [10]. When used to simulate outdoor environments, in-lab driving simulators can be used to test the usage of AR on roads [53]. In-lab simulators can also be used to test how other road users, such as pedestrians and cyclists, interact with autonomous vehicles [23, 38]. However, simulator sickness remains a large risk for in-lab driving simulators because a user's vestibular senses do not align with their visual senses when taking part in a simulation [4]. Researchers have tried to address this with methods such as aligning vehicular motion with VR content [7] or simulating movement [11, 22], but even the highest-end simulators replicate a fraction of the forces felt in normal on-road driving. [19]

2.2 XR On-Road Driving Simulators

XR technology enables simulating AV driving either by allowing users to move through completely virtual spaces with real vehicle dynamics or by overlaying virtual objects on top of real-time video footage of the surroundings to increase the immersion of pre-programmed interfaces or interactions [6, 16, 17, 21, 45, 46, 48, 58]. Recently, XR systems have been deployed on-road to take advantage of the realistic road environment and vehicle dynamics. The XR-OOM system developed by Goedicke et al. [16] employed an XR headset for drivers to drive through virtual, external obstacles in a parking lot. The MAXIM system developed by Yeo et al. utilized a virtual reality headset coupled with 360° cameras for subjects to experience an autonomous virtual vehicle situated in a live environment created from live streamed 360° videos [58]. Ghiurău et al. [15] showed a proof-of-concept headset-based XR driving experience revealing that original equipment manufacturers (OEMs) such as Volvo use such technology. Finally, McGill et al. [39] presented PassengXR, an open-source toolkit to create passenger XR experiences. While providing XR experiences, they did not compare their system to an in-lab simulator.

Although all the systems mentioned above track the vehicles' dynamics to accurately position virtual objects related to the participant, they do not natively support high-precision interaction staging based on the surroundings outside of the vehicle, which we refer to as *surrounding context* in this paper. In essence, previous approaches are mostly concerned with aligning virtual and physical motions, such as CarVR, but not the worlds themselves [21]. The authors of PassengXR introduced a hypothetical application that requires high-precision alignment between virtual and physical worlds, where passengers on a bus tour could view AR-style information overlaid on historic buildings outside the vehicle, but their demonstration was still carried out indoors without discussing how feasible it was to implement such application on-road [39]. Potential issues for designing such applications with existing platforms are two-fold. First, current systems make it challenging for designers to stage interactions, where they need to manually locate event trigger positions (in the bus tour example, buildings' coordinates either in GPS coordinates or relative coordinates in the vehicle frame) and program corresponding AR information boards to be at those precise coordinates. Second, the GPS-based tracking system may not provide enough accuracy for small-scale interaction, especially in cities where buildings can disrupt GPS signals. The authors for PassengXR commented that their "approach prioritizes perception of motion over location accuracy," which is not best suited for location-based XR experience [39]. These hardships limit the capability of these on-road systems to act as participant testing platforms compared to traditional in-lab simulators. Most autonomous driving studies require sufficient staging and surrounding context [8]. For example, driver-to-driver communications mostly happen at intersections, and pedestrian-vehicle interactions usually occur at crosswalks [24, 35, 37, 50]. Our work makes high prevision interaction staging possible while keeping the design process intuitive.

2.3 Platform Portability

Previous research has focused on how to replicate on-road scenarios in in-lab simulators, which is crucial when studying problems that are dangerous to experiment on the road, such as near-collision scenarios and passive rail level crossing [13, 36]. However, compromises are necessary to compensate for the lack of motion and sensory cues in in-lab simulators, and little research has been conducted to reduce the performance gap between on-the-road and in-simulators [18]. As such, merely pursuing statistical significance with in-lab simulators may result in overlooking issues of practical relevance in real-world contexts. [18].

Because in-lab and on-road driving simulation environments offer different strengths when it comes to control and realism of driving scenarios, it can be desirable to run the same study in both when possible— an approach we call twinning of studies — to understand how study results from different environments relate to each other. Hammel et al. [20] found that, when they replicated an on-road study in a fixed-based simulator, participants' eye-scanning behavioral patterns were similar, which demonstrates fix-based simulators' ecological validity. In a systematic review of validation studies featuring comparisons of driving simulation and on-road driving between 1977 and 2017, authors Wynne et al. [57] found only 44 validation studies comparing simulation to real driving. This is out of the 21,312 found by the same researchers to be English-language publications of original research having to do with driving simulation. Such studies are so rare that the 44 represent less than 0.25% of the published driving simulation research Wynne et al. [57]. They report that "There was little consistency in the dependent measures used to assess differences between the simulator and on-road drive...Of particular concern is the fact that only half of the driving simulators were found to be valid and some were valid for one measure but not others." They note that since policy, legislation, and training are built off of simulator studies, a better understanding of which aspects of simulated studies are likely to carry over to real road conditions, and which are not is critical [57].

Frequently, we believe, the lack of validation studies is due to the significant challenge of creating "twinned" studies in both environments. The advent of on-road mixed reality simulation [2, 16, 39, 58, 59] makes it possible to bridge the divide using software events in the real world. However, no system has yet ported the same study course, code, and event design from one environment to the other. By making it possible to port studies developed for in-lab simulators to be run on-road–what we call *platform portability*–we improve the ability for automotive researchers to extend their in-lab studies to the real road, and thereby improve the validity of simulation research. This would improve the ability of researchers to validate their simulation studies in on-road environments, as recommended by Wynne et al. [57]: "Ideally this would see authors report empirical validation evidence for their own simulator, and not relying on other simulators as support for validity. Even if modeled on a previously validated simulator, each set-up is unique and should be validated for those specifications."

3 SYSTEMS

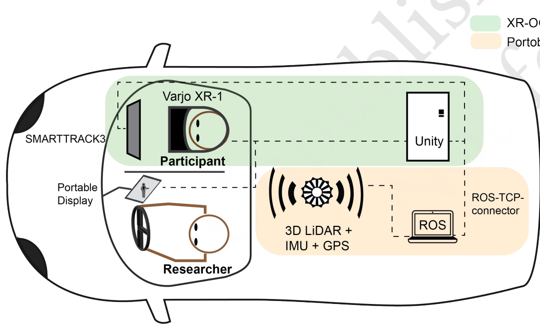

Figure 3: The Portobello system uses a LiDAR-based navigation stack to localize the vehicle's runtime position within a given map. The location information is transmitted to the Unity Desktop through ROS. The black dotted line indicates a virtual divider appearing only in the headset.

In this work, we present a study we developed meant to run on two driving simulator environments: a lab-based driving simulator and an on-road XR driving simulator equipped with our Portobello system. Here, we recap the features of both environments and introduce the key adjustments made to accommodate the Portobello system.

3.1 In-Lab Driving Simulator

Our in-lab fixed-base driving simulator features a modified Fiat 500 in front of three projector screens (see Figure 1). The projector screens cover participants' visual field when they sit in either the driver's seat or the passenger's. The three projectors are DLP-based and can produce an image with low latency on the projector screens. The projectors are connected to the computer over HDMI and use the TripleHead2Go to split one DisplayPort signal into the three outputs.

The vehicle is coated with non-reflective material to reduce the backscatter onto the projector screens, increasing the contrast of the projector screens. ButtKicker haptic transducers are installed under the front seats to provide realistic tactile feedback from road noise and the engine. The simulation software is run by an Alienware Area-51 R4 computer with two NVIDIA GTX1080 in SLI. The vehicle's side mirrors are small digital displays rendered by the same computer.

In our simulation software, the simulated vehicle is extended out of the Genivi Vehicle Simulator, which has been widely used in driving simulation studies [1]. The simulated vehicle uses standard Unity physics wheel colliders with a built-in engine simulation. To enable autonomous driving, we replace the original steering wheel input with a waypoint-based navigation system.

3.2 On-Road XR System

To enable cross-environment study deployment, we used Portobello with the XR-OOM system designed by Goedicke et al. [16]. In the original XR-OOM system, tracking and positioning of virtual objects are managed by a ZED 2 camera (for visual-SLAM) and the ART SMARTTRACK3[1] (for headset tracking within the vehicle) [16]. An onboard desktop running the Unity 3D game engine in version 2020.3.26f1 overlays virtual objects on top of the "passed-through" video of surroundings in the XR headset.

3.3 Portobello System

In the Portobello system, we used a LiDAR-based navigation system on the car roof driven by the Robot Operating System (ROS 1 Noetic) [44]; this replaces the XR-OOM's ZED 2 camera in front of the vehicle. The Portobello navigation system features an Ouster OS-1 3D LiDAR with a built-in Inertial Measurement Unit (IMU) and a ZED-F9R u-Blox GPS module. The communication between the navigation system and the Unity Desktop is managed through the ROS-TCP-Connector[2] provided by Unity [29].

Platform portability also drove augmentations to the XR-OOM software structure. Whereas XR-OOM used real-time visual SLAM to compute short-term vehicle odometry, the Portobello LiDAR system enables global vehicle localization within a given map. As a result, virtual objects' positions are no longer associated with the vehicle position directly as its children. Rather, the vehicle and virtual objects share a common parent — the world frame — which is introduced by a map of the environment. Instead of staging virtual events around the vehicle, as was done in XR-OOM (and in all the other XR-based driving simulation systems mentioned in [15]),

[1] https://ar-tracking.com/products/tracking-systems/smarttrack3/, accessed Jan 20, 2023
[2] https://github.com/Unity-Technologies/ROS-TCP-Connector, accessed Oct 20, 2022

Portobello can stage virtual events in a static shared map through which the vehicle drives. As Portobello replaces the car-centered reference frame with a map-based global frame, out-of-vehicle virtual objects remain fixed with respect to the map instead of to the arbitrary starting position of the vehicle. We detail the map generation process and staging process in Section 4.

At runtime, Portobello's LiDAR-based navigation system updates the position of a virtual vehicle in Unity. The relative position and orientation between the virtual vehicle and the participant's headset are managed by SMARTTRACK3. The virtual vehicle has the same shape as our research vehicle and is aligned with the research vehicle throughout the ride. To simulate proper depth ordering, the virtual vehicle is transparent with the alpha clipping option enabled. By acting as a cutout shader, the virtual vehicle occludes virtual objects outside. From the participants' view, the virtual objects are occluded by the real car they sit in. As we are rendering virtual events in XR over a large area (half of an island), we capped the headset's maximum rendering distance to 45 meters in order to improve motion parallax.

From a system design perspective, the computer running Portobello with ROS is a separate computer from the computer running XR-OOM with Unity. In essence, the communication between the ROS localization algorithm and Unity is achieved through the ROS transform package (TF), which is a hierarchical tree structure that tracks the relative position of multiple coordinate frames (map, LiDAR, etc.). These coordinate frames can be accessed from Unity as game objects. Isolating the system on a hardware level allows researchers to develop ROS and Unity in their own environments and for one driving simulator to be swapped out for another. Designers can focus on designing studies by placing objects in the course map in Unity, rather than being concerned with low-level ROS localization of objects. Another benefit of this practice is that the Portobello system consumes no computational resources in the original on-road platform at run time. The LiDAR-generated point-cloud map is rendered into Unity in the design phase. At runtime, the on-road platform computer imports the Unity map at start-up, and does not require the additional computational resources that would be needed to manage the point cloud map data. (This does mean that the Unity map might be missing physical features that change between the design phase and runtime.)

4 ENABLING PORTABILITY OF STUDY DESIGN USING COMMON MODELS

To run twinned studies across platforms, we must keep portability in mind during the design of the study. We outline the necessary components to guarantee equivalent performance in cross-environment twinned studies and detail our system pipeline to showcase the components' connections using our system.

4.1 Course Design

On-road simulators are limited by real-world road infrastructures. Based on the study focus, researchers should carefully consider test routes to ensure efficiency and reproducibility.

As our study focused on interactions at crosswalks, we chose the southern loop of Roosevelt Island as our study course. Most of the 0.9-mile long route is single-track and has a high density of crosswalks, 15 in total. The drive takes about 8 minutes. The route has no traffic lights and overlaps with two bus lines.

4.2 Map

A map (model of the study area) is the starting point of event design, and its precision and resolution greatly affect design complexity and quality. In cross-environment twinned studies, a map is a bridge between simulation and the real world, and it is also the shared common ground on every simulator platform.

To generate a high-quality map, we used the LiDAR-based navigation system to scan the entire test area. Specifically, we ran the real-time LiDAR-inertial odometry package (LIO-SAM) developed by Shan et al. [51] to create a true-to-scale point cloud mapping of the study area. We drove through the testing route multiple times to ensure loop closure. The resulting map is a monochromatic digital twin that includes over one million points and captures the test area in fine detail. Researchers can manually contextualize the point cloud map in Unity with colored assets and use the map as the background in in-lab simulators.

4.3 Event Design

Staging events along the planned course requires researchers to consider two major questions: where and when events occur. Staging is relatively easy for in-lab simulations, where agents' movement and speed profiles can be carefully controlled to guarantee timing and location. In this section, we discuss how we stage events using on-road simulators.

4.3.1 *Planned Events.*

Where? A one-to-one scaled map is necessary to plan events for on-road simulations because the vehicle will drive through the real world during the study. Any scaling or shifting on the map will cause significant errors at runtime. With a loaded digital twin in Unity, researchers can drag and drop virtual agents and objects just like they would for in-lab simulators.

When? Timing of events for on-road simulators can be controlled through the placement of collision-based triggers. Triggers in Unity are colliders that trigger events upon external contact. For example, invisible triggers can be placed at some distance x in front of a virtual traffic light. Once the vehicle collides with the trigger, the traffic light starts changing colors, and participants should react accordingly. The distance x governs the start of the interaction, which essentially affects the maximum response time for participants.

Of course, a vehicle on-road cannot "collide" with the virtual collider in simulation. In our system, the LiDAR-based navigation system synchronized the position of the real vehicle with a virtual vehicle in the digital twin through the *hdl_localization* and *ROS-TCP-Connector* packages in real-time [33]. (By employing the *hdl*$_l$*ocalization* algorithm, the offset between the vehicle's actual and estimated locations is maintained within a 0.2-meter range.) As the real vehicle drives through the world, the virtual vehicle simultaneously moves through the digital twin to trigger planned events. Note that this is the same virtual vehicle mentioned in section 3.2 for proper depth ordering.

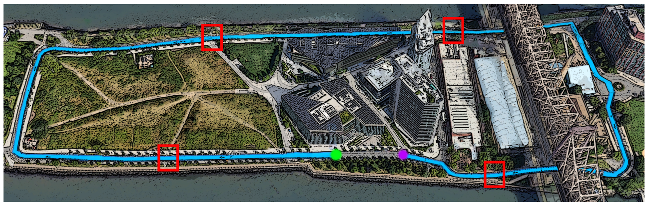

Figure 4: Our study area for on-road simulator on Roosevelt Island. The pre-determined test route is highlighted in blue. The crosswalks with staged interactions are highlighted in the red bounding boxes. The start and end locations are denoted by the green and purple dots, respectively.

4.3.2 Unplanned Events. We define events outside the simulation, which researchers have no control over, as *unplanned events*. In in-lab simulations, unplanned events are rare and typically caused by system failures or external interruptions. However, in on-road simulators, unplanned events are common and can even be valuable for ecological validity; they let researchers know if their findings are robust to real-world variation. Findings that are only true in the tightly controlled environment of a study have little practical application.

Nevertheless, researchers must factor in potential unplanned events during the design phase to ensure safety and preserve meaningful study results. Unplanned events come in different forms, from unexpected appearances of pedestrians to weather changes. For instance, in our study, we encountered the following unplanned events: real pedestrians and geese crossing the street, other vehicles passing the research vehicle from the bicycle lanes, and rain.

4.4 Platform Measures

Another aspect of *platform portability* is whether researchers can obtain the same set of measures from twinned studies. While measurements should be equally attainable across all simulation platforms, the characteristics of each simulator naturally encourage and discourage different sets of measures.

4.4.1 Behavioral Response. Behavioral responses refer to the participants' elicited behavior during the study. Examining behavioral responses is crucial when studying interactions between drivers, vehicles, and infrastructure [47]. For example, Jansen et al. [26] was interested in differences in participants' responses to different stimuli in automotive user interfaces. With appropriate sensors, collecting behavioral responses in in-lab and on-road simulators is possible.

4.4.2 Performance Response. Paas and Van Merriënboer [42] define performance as efficiency in completing tasks. We distinguish performance from behavioral responses based on the availability of ground truth. Researchers can collect performance responses during the study when participants are assigned tasks with general guidelines and standards. One example of performance response is the lateral vehicle position when the driver is distracted [34]. While extra sensors might be needed for on-road simulators to obtain vehicle-related performance measures (e.g., vehicle speed, acceleration, or trajectories are not easily attainable in on-road simulators as they are in in-lab simulators), we do not anticipate significant challenges in obtaining performance responses in both in-lab and on-road simulators.

4.4.3 Survey Response. Surveys can be conducted through different devices (pen and paper, tablets) in various formats (interviews, multiple choice, open-ended questions). In portable study design, it is important to consider the timing of the survey. One natural advantage of in-lab simulators over on-road simulators is the ability to pause at any point of the study and prompt participants with questions in situ [12]. On-road simulators cannot be paused easily, so surveys need to be planned so that participants can take them when it is safe to do so.

4.5 Additional Instrumentation

Detailed runtime recording of the environment is crucial for post-facto data analysis, particularly of unplanned events, for both on-road and in-lab simulators. Some measures need to be recorded differently in the different platforms and translated. For example, geo-location data from the on-road vehicle GPS must be correlated with the virtual world coordinates in the lab simulator. Head orientation and gaze direction obtained from the XR headset in the on-road simulator can be correlated with camera-tracked head-pose in the in-lab driving simulator.

5 TWINNING OF STUDIES

As proof of concept that we can run the same study design in the lab and on-road (twinning of studies), we conducted a within-subjects experiment with N=32 participants (under IRB#1806008105). We describe the cross-platform deployment of twinned studies and compare the differences in between. We counterbalanced the experiment conditions, where half of the participants experienced the study in the indoor simulator first and in the on-road simulator second, and the other 16 participants experienced the simulators in the reverse order.

We employed the in-lab and on-road driving simulators to run twinned Crosswalk Cooperation studies, which we adapted from Walch et al. [55]. In this previous study, Walch et al. used an in-lab driving simulator to evaluate the usability of a novel car UI and staged interactions around a driving loop. In this current work, we are not seeking to validate the results of the previously published study; we are not expecting or arguing that our study results would be the same. Rather, we are merely using this study design to evaluate the capacities and key influences of both systems.

5.1 Study Setup

5.1.1 Protocol. The experiment is a within-participants experiment design; each participant experiences two study sessions in counterbalanced order. In one session, participants experience the crosswalk cooperation study in the in-lab driving simulator. Afterwards, they fill out a post-session questionnaire, which collects information on their experience with the simulator. In the other session, the participants are escorted to the curbside and experience the crosswalk cooperation study in the on-road driving simulator. Afterwards, they fill out the same post-session questionnaire. Finally, they fill out a post-study questionnaire, which collects information on their perceived differences between the two simulators.

During each simulation session, participants are seated in the front passenger seat and informed that they will experience automated driving: The vehicle will stop at all crosswalks automatically. The vehicle will proceed autonomously when road conditions are clear (e.g. crosswalks without virtual pedestrians) and will ask for input from the participant via a smartphone interface, on how to proceed in unclear situations (e.g. virtual pedestrians walking towards crosswalks). Specifically, the vehicle will ask, "Is now safe to proceed?" on the phone interface while waiting at the crosswalk; when the participant feels it is safe to proceed, they press the "Proceed" button, and the car resumes its predefined route. If the participants decide it is not safe to proceed, they should wait until it is safe to press the button. The researcher driving the vehicle monitors when the participant presses the button and manually proceeds with the course if it is safe.

5.1.2 Differences in Simulator Setup. As much as possible, we maintained identical setups for the twinned studies. The key difference was that during the on-road simulation session, the researcher driving the car was mindful of the actual road conditions before proceeding with driving.

To maintain the narrative that the vehicle was driving autonomously, the researcher driving during the on-road simulation was masked by a black divider in the video pass-through headset (shown in Figure 3). A similar black divider was also installed in the in-lab simulator to maintain setup parity. During pilot sessions, we found it difficult to disguise the on-road vehicle as an AV due to differences in sound profile. In complex on-road conditions, the sound of pressing the pedal and rotating the steering wheel broke the illusion quickly. Thus, we decided to inform the participants of the divider's purpose and that there was a real researcher in the car with them operating the vehicle.

While a virtual map environment is required in in-lab simulation, the point cloud map is not required for on-road simulation due to the benefit of the video-pass-through headset. Therefore, after the event staging phase, we disabled the point cloud rendering in Unity to save computation power.

5.1.3 Scenarios. We recreated four scenarios from the original Crosswalk Cooperation study by Walch et al. [55]. Each scenario was engineered so that pedestrian interactions would only happen at crosswalks. In each scenario, virtual pedestrians interacted with each other on the sidewalks near the crosswalks. In half of the scenarios, one pedestrian walked to the stop sign and crossed the street after giving clear body language signals that they intended to cross (looking left and right). In the other half of the scenarios, the pedestrian stopped at or walked past the crosswalk. In our version of the Crosswalk Cooperation study, we constrained our study area to the southern loop of Roosevelt Island.

5.2 Participants

Out of our 32 participants ranging from 20 to 47 years old (age: M = 27.38 ± 5.92), 19 participants identified themselves as male, 11 as female, and two as non-binary. Six participants had experience with AV simulations and/or AV research, and five had experience with commercialized AVs (Tesla, demos at car shows). The others had little experience with AVs.

5.3 Study Measures

After each session, participants filled out a post-session questionnaire, which collected information on their experience with the simulator. After completing both sessions and corresponding questionnaires, participants filled out a post-study questionnaire, which collected information on their perceived differences between the two simulators.

We also recorded video and audio for all sessions run in both in-lab and on-road driving simulations to investigate participants' behavioral responses. For the in-lab driving simulator, a go-pro camera was pointed towards the participant to record their upper body. The simulated virtual environment was recorded using screen recording software. For the on-road driving simulator, similar to the setup in the in-lab simulator, a camera is mounted in the glove compartment to record the participant's upper body. An additional camera is mounted near the rear mirror facing forward to record the road condition ahead. The participant's XR view (video pass-through with overlay) is also recorded using the Varjo Base software.

6 RESULTS AND ANALYSIS

We analyzed the video footage and the questionnaires from both sessions with the goal of investigating the influences of both driving simulators on the user study experience and understanding what results from one platform predict for results on the other. Our evaluation of Portobello is based on being able to run and gather comparable results from the studies on both the in-lab and on-road platforms and not on the originality, validity, or significance of the study itself.

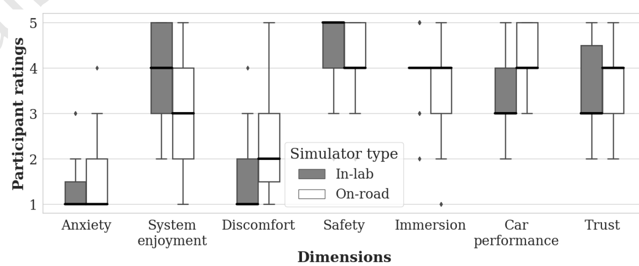

Figure 5: Participants' ratings of each simulator on a five-point Likert scale across seven distinct dimensions where 1=low and 5=high.

6.1 Study Results

6.1.1 Measures. To compare the overall participant measures in the study on both simulators, we asked the participants (N = 32) to rate their feelings of anxiety, safety, and trust on a 5-point Likert scale for each of the platforms. We designed the questions based on the questionnaire used in Walch et al. [55]'s original Crosswalk Cooperation study. We ran a Bayesian factor analysis on the captured measures with the null hypothesis that there is no difference between the platforms [31] with R version 4.3.2 and the BayesFactor package [40] using Jeffreys-Zellner-Siow (JZS) priors, i.e., the default non-informative Jeffreys prior. Interpretations were made according to Jeffreys [28]. All packages were up to date in November 2023. Since we have limited data, and we are not interested in how different variables (ratings) interact with each other, we chose to report single variate analysis over multivariate analysis. Nonetheless, our claims hold under multivariate analysis as well. We will provide R script for both analyses.

One participant left multiple answers empty, so we dropped their results in the following analysis.

Anxiety. More participants reported reduced anxiety with the in-lab simulator (M = 1.32 ± 0.60) than with the outdoor simulator (M = 1.45 ± 0.72). We found *moderate* evidence (BF = 0.32) in favor of the null model, suggesting that there is no significant difference in the anxiety generated by the simulators.

Safety. Participants consider the in-lab simulator (M = 4.58 ± 0.76) safer than the on-road simulator (M = 4.16 ± 0.86). We found *moderate* evidence (BF = 3.53) against the null model, suggesting a moderate difference in favor of the in-lab simulator.

This may have been because the in-lab simulator's roads did not involve any real vehicles or people, e.g., P6 explained that "... there were more actual obstacles...to take into account [in the outdoor simulator] whereas the in-lab [simulator] had a preset number."

Trust. Participants reported their trust in the simulated autonomous driving to be higher in the on-road simulator (M = 3.71 ± 0.86) than in the in-lab simulator (M = 3.52 ± 1.06). We found *moderate* evidence (BF = 0.31) in favor of the null model, suggesting there is no significant difference in trust.

The preference for the on-road simulator may have been influenced by people's perception of each vehicle's performance. At the same time, this may have been a breakdown in face validity. For example, P14 said, "One of the survey questions asked 'how much do you trust this car' and I think I forgot to pretend that this was an AI driving the car when answering that..." Participants were informed during the on-road simulator session that a researcher would be driving the car; their trust rating might have been an indicator that they trusted the driver to obey local traffic laws, rather than an indicator of their trust in the simulated autonomous driving, as we had intended.

6.1.2 Cooperation Behavior. From recorded videos, we analyzed 32 participants' cooperative behaviors with the vehicle at crosswalks. Video recordings from the same session were synchronized before the analysis. One researcher watched recordings for each participant and labeled their behaviors in terms of the timing of cooperation behaviors for each scenario. The researcher then noted down the behavioral changes, if any, for each participant between different simulators.

15 participants cooperated with the vehicle perfectly in both simulation platforms in all scenarios, where they waited until the pedestrians fully crossed the street or waited for clear non-crossing signals before instructing the vehicle to proceed. During the first crossing scenario, 11 participants instructed the vehicle to proceed while the pedestrian was about to cross; this led to six virtual collisions. One of these participants made the vehicle run over the same virtual pedestrian in both in-lab and on-road simulators. Three participants did not wait for any virtual pedestrian to cross in both in-lab and on-road simulators. Six participants who had cooperated perfectly in the first session made different decisions in the second session; they chose not to wait for the crossing participants when they believed it was safe, and one of them ended up running over the virtual pedestrian. One participant who ignored both crossing pedestrians in in-lab simulators waited for one of the crossing pedestrians in the on-road simulator. Overall, 14 participants made different decisions in the second session from the first session.

We feel compelled to point out that the fact that participants made poor crossing decisions is not a sign that the study or the driving simulation platforms were designed poorly; instead, it is precisely these sorts of outcomes that indicate the necessity for simulation platforms that enable studies with virtual pedestrians to be conducted prior to putting real pedestrians in harm's way. We do not expect this means that participants would run over real people in subsequent tests with real cars, but this does point out that participants are aware that they are not exposed to real danger in driving simulators [25]; impatience and lack of conscientiousness amongst some portion of the population are factors that any cooperative autonomous driving system would have to account for.

6.1.3 Ordering Effect. While we counterbalanced our study, we noticed some differences in cooperation behavior that may be attributed to the ordering of simulators. Of the six participants who made mistakes in the first crossing scenario, four participants were experiencing the on-road driving simulator. We hypothesize that the on-road driving simulator is more overwhelming than the in-lab simulator to familiarize the participants with the study setup. P28 mentioned in their post-study questionnaire that "Visual noise in outdoor sim [made] task completion more difficult but [was] more realistic in that regard." It is worth mentioning that the three participants who made the mistake in the first crossing scenario had limited (e.g., they had only their learner's permits) to no driving experience.

6.2 Simulation Evaluation Results

6.2.1 Simulator Measures. To compare the overall experience between the two simulators, we also asked participants to rate their feelings about car performance, system enjoyment, discomfort, and immersion on a 5-point Likert scale. These questions were based on questionnaires used during the validation process of Goedicke et al. [16]'s XR-OOM system. We again ran a Bayesian factor analysis on the captured measures with the null hypothesis that there is no difference between the platforms [31].

Car Performance. Participants considered the autonomous vehicle in the on-road simulator (M = 4.35 ± 0.71) to perform better than in the in-lab simulator (M = 3.42 ± 0.96). We found *extreme* evidence (BF = *1.67e+04*) against the null model, suggesting a significant difference in favor of the on-road simulator. Participants felt that the in-lab car simulator did not appear to drive smoothly and stopped rather abruptly at times and thus thought that the driving felt more natural in the on-road simulator.

System Enjoyment. More participants reported increased levels of system enjoyment with the in-lab simulator (M = 3.84 ± 0.97) than with the outdoor simulator (M = 3.03 ± 0.98). We found *extreme* evidence (BF = 225.00) against the null model, suggesting a significant difference in favor of the in-lab simulator. Participants seemed to have preferred the in-lab simulator because of the graphics quality and comfort. For example, when asked to describe their experience with both simulators in the post-study questionnaire, P14 said, "The pedestrians seemed to "appear out of nowhere" in the [on-road simulator], whereas it seemed like they were always part of the scenery in the [in-lab] simulator (i.e., came into view naturally in the simulator). ... [The] turns in the [in-lab simulator’s] road felt unnatural/like the scenery was clicked and dragged in front of my eyes, instead of me moving through the scenery."

Discomfort. Participants reported less discomfort with the in-lab simulator ($M = 1.71 \pm 0.86$) than the on-road simulator ($M = 2.35 \pm 1.14$). We found *anecdotal* evidence (*BF = 2.24*) against the null model, suggesting a mild difference in favor of the in-lab simulator. For the in-lab simulator, discomfort mainly arose from the unrealistic vehicle dynamics. For the on-road simulator, many comments were related specifically to the XR headset ($n = 11$), which people found heavy and uncomfortable. The occasional misalignment of virtual objects caused by bumps in the road also induced a considerable amount of motion discomfort. For example, P13 said, "...both [simulators] cause some discomfort, but I think the outdoor one is more uncomfortable due to the [pass-through] being very [shaky] and more motion-sickness-inducing."

Immersion. As shown in Figure 5, participants considered the in-lab simulator ($M = 4.00 \pm 0.77$) to be more immersive than the on-road simulator ($M = 3.74 \pm 1.06$). We found *anecdotal* evidence (BF = 0.0.40) in favor of the null model. However, in the post-study questionnaire, when both simulators were presented on the same Likert scale, the numbers of participants in favor of either simulator were the same. Figure 6 shows that 14 participants thought the in-lab simulator was more immersive, and 14 thought the on-road simulator was more immersive. Three participants considered both simulators equally immersive.

Two participants reported verbally during the on-road session that it was difficult to distinguish virtual from real pedestrians.

Many participants reported difficulty with the weight and technical maturity of the XR headset in the outdoor simulator, which impeded their attention and may have contributed to its lower immersion rating. P8 said, "headset jitter made visuals blurry, which impaired decision-making/attention." P2 said, "[Putting] something on my head is so uncomfortable. I couldn’t [focus on] the view or be relaxed. The [in-lab] simulator is not so real[,] but I could be so relaxed."

6.2.2 *In simulator behavior.*

We weighed the observed and reported behavior of participants to assess differences in behavior introduced by each platform.

Natural Head Motions. We noticed that some participants’ head pose motion patterns were different when experiencing the two simulation platforms. When sitting in the in-lab static simulator, participants tilted their heads mostly during staged events to track the motion of virtual pedestrians. For the rest of the ride, they faced forward. However, for the on-road simulator, the participants’ head motions were more varied, and participants naturally looked to view the surroundings more often.

Decision-making. For the crosswalk cooperation study, the key measure is the participant’s decision-making around whether the vehicle should go or not go. Participants reflected that the level of complexity and severity in decision-making was greater in the on-road simulator. The staged interactions are identical in both sessions, but researchers have less control over surrounding factors during the on-road sessions. Other road users, including real pedestrians, other vehicles, and geese native to Roosevelt Island complicated the staged scenarios.

P5 said, "I think that in the on-road simulator, I was a little more nervous because real people were on the street. The in-lab simulator does not deal with real people, so any mistake I make does not have as much weight." P28 said, "Outdoor provides [a] more generally immersive feeling and is the only one in which I can realistically feel unsafe, which is a positive in terms of validity. Visual noise in outdoor sim makes task completion more difficult but is more realistic in that regard." P31 said, "I felt anxious after saying to proceed and wondering how the pedestrians would move afterwards." In contrast, P11 reflected: "Perhaps due to its nature of being [an in-lab] simulator, I felt at ease at all times." P20 reflected: "the [polished nature] of the [in-lab] simulator makes the experience [feel] more entertaining/performative rather than realistic. ... The [on-road] simulator feels more like a functional approximation of an actual autonomous driving car, while the indoor simulator feels more like a fun, polished experience." P21 reflected: "I felt much safer/less anxious in the indoor simulator which also probably means it was less realistic." P29 reflected: "[the] indoor [simulator] seems to have lower stakes, even though it was the same virtual people.

7 TECHNICAL VALIDATION

Since the Portobello system is an infrastructure meant to be used in conjunction with existing platforms, the technical performance largely depends on the system on top of Portobello. Therefore, we investigated the change in performance of the on-road platform after adapting Portobello. Rendering a total of 12 virtual pedestrians, the headset runs at 60 FPS consistently with a display latency of around 35ms, which is on par with the original XR-OOM system. The localization frequency is 10 Hz, limited by the 10Hz LiDAR. The fact that there is no change in performance is expected since the Portobello system operates on a computer separated from the original XR-OOM system.

8 DISCUSSION

8.1 LiDAR-based vs. GPS/IMU-based systems

Our LiDAR-based Portobello system resolves the two challenges posed by traditional GPS/IMU-based systems in surrounding context-based interaction staging: localization accuracy and design hardships. LiDAR-based localization systems work reliably in cities where buildings serve as landmarks instead of GPS signal blockers. A point cloud 3D map of the environment generated from LiDAR-based SLAM algorithm saves designers from staging using hardcoded coordinates, and simplifies the design process to drag-and-drop within the map. Lastly, we want to point out that the sensors are not mutually exclusive. We can fuse in GPS data as an additional data source into the LiDAR-based algorithm if necessary.

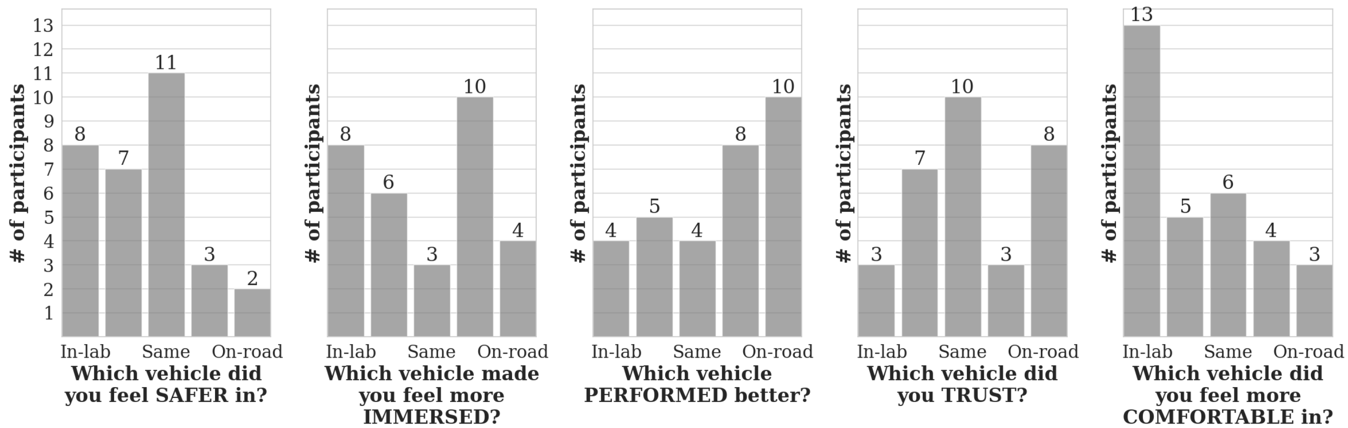

Figure 6: After experiencing both simulators, participants also directly compared the in-lab simulator and the on-road simulator on the same Likert scale. Participants' responses are shown as histograms.

8.2 Platform Portability Challenges

The primary goal of this research effort was to establish a proof-of-concept demonstration of *platform portability* through our Portobello system. *Platform portability* is necessary to run twinned studies across different platforms, which is desirable because the in-lab simulator can help establish causal differences across experimental conditions using well-controlled studies, and the on-road simulator can help validate the ecological validity of such study results when the same study conditions are moved into the less-controlled environment of the real world. By incorporating robotics mapping, sensing, and localization capability in the Portobello system, we can set up twinned studies to run in different environments. We believe that this is of relevance to any of the simulators discussed in Section 2.2, which could be deployed on top of Portobello.

8.2.1 Randomness in the Wild. For on-road simulators, randomness persists throughout the entire study. For example, during mapping, we generated a snapshot of the test area. While major landmarks such as buildings and land topology will not change significantly over time, the map also captures transient objects (e.g., parked cars). Such randomness may cause inaccuracy in real-time localization.

During the study, unplanned events were the most salient form of randomness. Unplanned occurrences and interruptions from the real world may increase immersion. P19 reported that "the extra people and cars in the outdoor study made the experience feel more immersive and interesting." However, randomness also brings concerns regarding study reproducibility. We tried to eliminate the co-occurrence of planned and unplanned events by staging events in less populated areas. In general, we encourage researchers to plan for all possible unplanned events during the study design phase.

8.2.2 Event Timing. One major challenge we faced with the on-road simulator is the trigger and timing of staged events, and we foresee such a problem persisting in future similar studies. For in-lab simulators, vehicle speed and travel distances can be coded in detail. For on-road simulators, it can be difficult to maintain the same speed curve as in-lab simulators due to obstacles and unplanned events. In our study, we expect the vehicle to stop at the crosswalk simultaneously as a virtual pedestrian reaches the stop sign on the sidewalk. The researcher who operates the car has access to a mini-display monitoring the location of the virtual pedestrians and adjusting vehicle speed accordingly. However, we have noticed that participants made different decisions across platforms due to event timing differences.

The timing misalignment between the simulators is the natural consequence of the intentional difference between running studies in a controlled environment (the lab) and an uncontrolled environment (the real world). We are arguing that it is desirable to run both kinds of studies and that it is easier to do this if *platform portability* exists. The in-lab simulator is more suitable for quantitative analysis, and the on-road simulator is more suitable for qualitative analysis of the factors that complicate the outcomes learned from the more controlled in-lab simulator.

8.3 Platform Effects

Our study results were intended to help us understand *whether and how* our twinned studies were the same across the two platforms and to help us understand the differences across the platforms, a comparison made possible by the Portobello system. For research purposes, it would be best if the study results between the two platforms were similar (i.e., that the platform effects were negligible), or at least that the results were biased consistently across the platforms (i.e., that the platform effects were predictable).

From the study results, we can see that around half of the participants made similar decisions in both the in-lab simulator and the on-road system. Notably, one participant made the same mistake in each simulator. The fact that some participants made different decisions in different simulators indicates that participants' decision-making was not affected by their existing knowledge about the study. Even if they were aware of the pedestrians' behaviors from the first session, they took into account the event timing difference and made the most appropriate decision at the moment.

Some notable differences between the platforms were centered around the participant’s decision-making behavior and their resulting trust in the autonomous driving system. In some ways, there were indications that the on-road simulator might have failed to maintain face validity for at least one of the participants; their trust rating might have shown their trust in the research driver rather than their trust in the simulated autonomous driving, as we had intended. On the other hand, many participants reported greater weight in the decision-making around whether the vehicle should go or not go in the on-road simulator, a sign that face validity is higher in the real-world environment.

While our participants favored the experience of the in-lab simulation platform on a whole, most of their complaints pertained to aspects of the XR system–the weight of the headset, the jitter in the display–which are likely to improve with advancing technology. It seems like the naturalism of the on-road environment is more likely to yield naturalistic behaviors than the in-vehicle environment, and hence this platform can help industrial and academic researchers better understand how people will engage with autonomous vehicle technologies in the real world.

8.4 Limitations and Future Work

Some limitations in the study results are inherent to a driving simulation study. In this section, we focus on limitations in the design and execution of studies using the XR driving simulation system augmented by Portobello that should be accounted for and discuss future developments that could improve such systems.

8.4.1 Real-time adjustment of Depth Ordering. Although we render the occlusion of virtual objects caused by the research vehicle (e.g. we do not render the pedestrians over the front pillar of the car), the current system does not provide the same occlusion for runtime dynamic objects. If a bus drove between the research vehicle and the location where the virtual pedestrians were supposed to appear, for example, participants would see the virtual pedestrians in front of the bus. To correct the depth order, future systems could use real-time LiDAR scans of the environment.

8.4.2 Pedestrian Appearance. We capped the maximum rendering distance for on-road virtual objects for technical reasons; distant virtual objects are less salient and require better alignment between the in-vehicle and out-of-vehicle reference frames to be placed believably in the mixed-reality view. The artifice of having pedestrians suddenly appear, however, may affect study results. One participant (P24) said, "I think being able to see the passengers from further away in the in-lab simulator made a big difference because, by the time I got to the intersection, it was easier to anticipate their movements." Future technology could improve the motion parallax issues, enabling longer rendering distances and smoother transitions when virtual objects approach the rendering threshold.

8.4.3 Headset Discomfort. Many participants complained about the bulkiness and narrow field-of-view of the headset. This platform-level discomfort was pronounced enough that it drowned out our ability to measure experiential aspects (system enjoyment, discomfort) of the autonomous driving scenario. While the weight and limitations of the XR headsets were beyond our control, we believe that anticipated advancements in XR headset technology are necessary to use these systems in experiments wherein the experiential aspects of automated driving are critical.

8.4.4 Simulating Autonomous Driving on Road. In our current study, we informed the participants that there was an actual driver behind the scenes in the on-road simulator. The driver’s maneuver sound easily breaks the AV illusion for participants who have previous experience with AV. Future research can benefit from disguising the driver by playing the recorded AV sound profile during acceleration and deceleration to cover the driver’s maneuver sound.

8.4.5 Consistent Driver Performance. While the same researcher operated the vehicle throughout the study, and attempted to maintain a consistent driving style from one run to the next, there were natural variations in the driving, in part in response to uncontrolled environmental factors. While it is not feasible nor desirable to force the driver to operate the vehicle the same way for all sessions on the road, future systems should collect data to enable later analysis of variance.

9 CONCLUSION

Driving simulations can be used to create scenarios for driving interactions, which enable researchers to better understand how people will behave and respond to future driving scenarios. In this work, we presented the Portobello system, an on-road driving simulation infrastructure that enables *platform portability*. By advancing the capabilities of driving simulators, we can better anticipate what aspects of driving interaction will work well or poorly.

This paper outlines the first-ever deployment of twinned studies across in-lab and on-road simulators. We found that participants preferred the experience of the in-lab simulator but displayed more natural head movements in the on-road simulator; they also reported that the decisions made in the on-road system carried more weight. Based on our findings, we suggest researchers working in driving simulations also take the twinning of studies approach: they should first run studies within a controlled, in-lab environment to collect statistical measures and form hypotheses and then port their studies to a less-controlled, on-road simulator and test their hypotheses in a more complex, realistic environment.

This experiment looking at the platform-driven influences on study outcomes demonstrates the utility of *platform portability*, as the same study design was able to be run both in-lab and on-road. This was made possible by the Portobello system’s common model and vehicle localization; using robotics mapping and localization technology, we were able to capture surrounding environments for study and event staging for our on-road simulator. We anticipate that Portobello will advance the state of open-source and accessible driving simulation by extending the reach and translational capabilities of VR and XR driving simulation systems and thus, in turn, enable wider-scale development and testing of safe driving systems.

OPEN SCIENCE

The source code for Portobello has been made publicly available. It can be accessed via the following link:
https://github.com/FAR-Lab/Portobello.git.

ACKNOWLEDGMENTS

This research was made possible by sponsorship by Toyota Research Institute and Woven Planet. Studies were conducted under Cornell Tech's IRB Protocol #IRB0008105. We would like to express our sincere gratitude to Dr. Hiroshi Yasuda for his expert guidance and insightful suggestions throughout the development of this research. This work was also supported by a fellowship of the German Academic Exchange Service (DAAD).

REFERENCES

[1] GENIVI Alliance. 2021. Genivi Vehicle Simulator. https://github.com/GENIVI/genivi-vehicle-simulator.

[2] Sonia Baltodano, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju. 2015. The RRADS platform: a real road autonomous driving simulator. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, New York, NY, USA, 281–288.

[3] Karlin Bark, Cuong Tran, Kikuo Fujimura, and Victor Ng-Thow-Hing. 2014. Personal navi: Benefits of an augmented reality navigational aid using a see-thru 3d volumetric hud. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, New York, NY, USA, 1–8.

[4] Giovanni Bertolini and Dominik Straumann. 2016. Moving in a moving world: a review on vestibular motion sickness. *Frontiers in neurology* 7 (2016), 14.

[5] David Bethge, Thomas Kosch, Tobias Grosse-Puppendahl, Lewis L. Chuang, Mohamed Kari, Alexander Jagaciak, and Albrecht Schmidt. 2021. VEmotion: Using Driving Context for Indirect Emotion Prediction in Real-Time. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 638–651. https://doi.org/10.1145/3472749.3474775

[6] Björn Blissing, Fredrik Bruzelius, and Johan Ölvander. October 23-25, 2013. Augmented and mixed reality as a tool for evaluation of vehicle active safety systems. In *RSS2013 Road Safety and Simulation-International Conference*. Aracne, Rome, Italy, 7 pages.

[7] Hyung-jun Cho and Gerard J. Kim. 2020. RoadVR: Mitigating the Effect of Vection and Sickness by Distortion of Pathways for In-Car Virtual Reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology (VRST '20)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3385956.3422115

[8] Mark Colley, Julian Britten, Simon Demharter, Tolga Hisir, and Enrico Rukzio. 2022. Feedback Strategies for Crowded Intersections in Automated Traffic — A Desirable Future?. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 243–252. https://doi.org/10.1145/3543174.3545255

[9] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 155, 11 pages. https://doi.org/10.1145/3411764.3445351

[10] Mark Colley, Pascal Jansen, Enrico Rukzio, and Jan Gugenheimer. 2021. Swivr-car-seat: Exploring vehicle motion effects on interaction quality in virtual reality automated driving using a motorized swivel seat. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–26.

[11] Mark Colley, Pascal Jansen, Enrico Rukzio, and Jan Gugenheimer. 2022. SwiVR-Car-Seat: Exploring Vehicle Motion Effects on Interaction Quality in Virtual Reality Automated Driving Using a Motorized Swivel Seat. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 150 (dec 2022), 26 pages. https://doi.org/10.1145/3494968

[12] Joost De Winter, Peter M van Leeuwen, Riender Happee, et al. 2012. Advantages and disadvantages of driving simulators: A discussion. In *Proceedings of measuring behavior*, Vol. 2012. 8th.

[13] Sebastien Demmel, Gregoire S Larue, Andry Rakotonirainy, and Dominique Gruyer. 2011. Using cooperative vehicle systems to collect near collision incidents events: a simulation study. In *2nd International Workshop on Traffic Data Collection & its Standardisation*.

[14] Lorin Dole and Wendy Ju. 2019. Face and Ecological Validity in Simulations: Lessons from Search-and-Rescue HRI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300681

[15] Florin-Timotei Ghiurău, Mehmet Aydın Baytaş, and Casper Wickman. 2020. ARCAR: On-Road Driving in Mixed Reality by Volvo Cars. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 62–64. https://doi.org/10.1145/3379350.3416186

[16] David Goedicke, Alexandra W.D. Bremers, Sam Lee, Fanjun Bu, Hiroshi Yasuda, and Wendy Ju. 2022. XR-OOM: MiXed Reality Driving Simulation with Real Cars for Research and Design. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 107, 13 pages. https://doi.org/10.1145/3491102.3517704

[17] David Goedicke, Jamy Li, Vanessa Evers, and Wendy Ju. 2018. Vr-oom: Virtual reality on-road driving simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–11.

[18] Paul Allan Green. 2005. *How driving simulator data quality can be improved.* University of Michigan Transportation Research Institute, Human Factors Division.

[19] Jeff Greenberg, Bruce Artz, and Larry Cathey. 2003. The effect of lateral motion cues during simulated driving. *Proceedings of DSC North America* (2003), 8–10.

[20] Kim R Hammel, Donald L Fisher, and Anuj K Pradhan. 2002. Verbal and spatial loading effects on eye movements in driving simulators: A comparison to real world driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 46. SAGE Publications, Los Angeles, CA, USA, 2174–2178.

[21] Philipp Hock, Sebastian Benedikter, Jan Gugenheimer, and Enrico Rukzio. 2017. CarVR: Enabling In-Car Virtual Reality Entertainment. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4034–4044. https://doi.org/10.1145/3025453.3025665

[22] Philipp Hock, Mark Colley, Ali Askari, Tobias Wagner, Martin Baumann, and Enrico Rukzio. 2022. Introducing VAMPIRE – Using Kinaesthetic Feedback in Virtual Reality for Automated Driving Experiments. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 204–214. https://doi.org/10.1145/3543174.3545252

[23] Ming Hou, Karthik Mahadevan, Sowmya Somanath, Ehud Sharlin, and Lora Oehlberg. 2020. Autonomous Vehicle-Cyclist Interaction: Peril and Promise. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376884

[24] Maura Houtenbos. 2009. Interacting drivers at intersections: what can make them more safe and more efficient. In *Proceedings of the 21st International Technical Conference on the Enhance Safety Vehicles*. Citeseer, Citeseer.

[25] Samantha L Jamson and A Hamish Jamson. 2010. The validity of a low-cost simulator for the assessment of the effects of in-vehicle information systems. *Safety Science* 48, 10 (2010), 1477–1483.

[26] Pascal Jansen, Julian Britten, Thilo Segschneider, Mark Colley, and Enrico Rukzio. 2023. AutoVis: Enabling Mixed-Immersive Analysis of Automotive User Interface Interaction Studies. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580760 conditionally accepted.

[27] Pascal Jansen, Mark Colley, and Enrico Rukzio. 2022. A Design Space for Human Sensor and Actuator Focused In-Vehicle Interaction Based on a Systematic Literature Review. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 56 (jul 2022), 51 pages. https://doi.org/10.1145/3534617

[28] Harold Jeffreys. 1998. *The theory of probability.* OUP Oxford.

[29] Arthur Juliani, Vincent-Pierre Berges, Ervin Teng, Andrew Cohen, Jonathan Harper, Chris Elion, Chris Goy, Yuan Gao, Hunter Henry, Marwan Mattar, et al. 2018. Unity: A general platform for intelligent agents. *arXiv preprint arXiv:1809.02627* (2018).

[30] Mohamed Kari, Tobias Grosse-Puppendahl, Alexander Jagaciak, David Bethge, Reinhard Schütte, and Christian Holz. 2021. SoundsRide: Affordance-Synchronized Music Mixing for In-Car Audio Augmented Reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 118–133. https://doi.org/10.1145/3472749.3474739

[31] Robert E Kass and Adrian E Raftery. 1995. Bayes factors. *Journal of the American statistical association* 90, 430 (1995), 773–795.

[32] SeungJun Kim and Anind K Dey. 2009. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 133–142.

[33] Kenji Koide, Jun Miura, and Emanuele Menegatti. 2019. A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems* 16, 2 (2019), 1729881419841532.

[34] Georgios K Kountouriotis and Natasha Merat. 2016. Leading to distraction: Driver distraction, lead car, and road environment. *Accident Analysis & Prevention* 89 (2016), 22–30.

[35] Sven Krome, David Goedicke, Thomas J Matarazzo, Zimeng Zhu, Zhenwei Zhang, JD Zamfirescu-Pereira, and Wendy Ju. 2019. How people experience autonomous

intersections: taking a first-person perspective. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, New York, NY, USA, 275–283.

[36] Grégoire S Larue, Christian Wullems, Michelle Sheldrake, and Andry Rakotonirainy. 2018. Validation of a driving simulator study on driver behavior at passive rail level crossings. *Human factors* 60, 6 (2018), 743–754.

[37] Markus Lemmer, Stefan Schwab, and Sören Hohmann. 2020. Driver interaction at intersections: A hybrid dynamic game based model. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, IEEE, New York, NY, USA, 2269–2276.

[38] Karthik Mahadevan, Elaheh Sanoubari, Sowmya Somanath, James E. Young, and Ehud Sharlin. 2019. AV-Pedestrian Interaction Design Using a Pedestrian Mixed Traffic Simulator. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 475–486. https://doi.org/10.1145/3322276.3322328

[39] Mark McGill, Graham Wilson, Daniel Medeiros, and Stephen Anthony Brewster. 2022. PassengXR: A Low Cost Platform for Any-Car, Multi-User, Motion-Based Passenger XR Experiences. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 2, 15 pages. https://doi.org/10.1145/3526113.3545657

[40] Richard Morey. 2022. *Richarddmorey/BayesFactor*. Retrieved 6 September, 2022 from https://github.com/richarddmorey/BayesFactor

[41] Wolfgang Narzt, Gustav Pomberger, Alois Ferscha, Dieter Kolb, Reiner Müller, Jan Wieghardt, Horst Hörtner, and Christopher Lindinger. 2006. Augmented reality navigation systems. *Universal Access in the Information Society* 4 (2006), 177–187.

[42] Fred GWC Paas and Jeroen JG Van Merriënboer. 1993. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors* 35, 4 (1993), 737–743.

[43] Maria Pinto, Viola Cavallo, and Théophile Ohlmann. 2008. The development of driving simulators: Toward a multisensory solution. *Le travail humain* 71, 1 (2008), 62–95.

[44] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, Japan, 5.

[45] Michael Quinlan, Tsz-Chiu Au, Jesse Zhu, Nicolae Stiurca, and Peter Stone. 2010. Bringing simulation to life: A mixed reality autonomous intersection. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, New York, NY, USA, 6083–6088. https://doi.org/10.1109/IROS.2010.5651993

[46] Andreas Riegler, Andreas Riener, and Clemens Holzmann. 2020. A research agenda for mixed reality in automated vehicles. In *19th International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, USA, 119–131.

[47] Daniele Ruscio, Luca Bascetta, Alessandro Gabrielli, Matteo Matteucci, Dedy Ariansyah, Monica Bordegoni, Giandomenico Caruso, and Lorenzo Mussone. 2017. Collection and comparison of driver/passenger physiologic and behavioural data in simulation and on-road driving. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, New York, NY, USA, 403–408.

[48] Hiroaki Sawano and Minoru Okada. 2005. A car-navigation system based on augmented reality. In *ACM SIGGRAPH 2005 Sketches*. ACM, New York, NY, USA, 119–es.

[49] Mark C Schall Jr, Michelle L Rusch, John D Lee, Jeffrey D Dawson, Geb Thomas, Nazan Aksan, and Matthew Rizzo. 2013. Augmented reality cues and elderly driver hazard perception. *Human factors* 55, 3 (2013), 643–658.

[50] Friederike Schneemann and Irene Gohl. 2016. Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. In *2016 IEEE intelligent vehicles symposium (IV)*. IEEE, IEEE, New York, NY, USA, 38–43.

[51] Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Rus Daniela. 2020. LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, IEEE, New York, NY, USA, 5135–5142.

[52] Yoshisuke Tateyama, Hiroki Yamada, Junpei Noyori, Yukihiro Mori, Keiichi Yamamoto, Tetsuro Ogi, Hidekazu Nishimura, Noriyasu Kitamura, and Harumi Yashiro. 2010. Observation of drivers' behavior at narrow roads using immersive car driving simulator. In *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '10)*. Association for Computing Machinery, New York, NY, USA, 391–396. https://doi.org/10.1145/1900179.1900263

[53] Marcus Tonnis, Christian Sandor, Gudrun Klinker, Christian Lange, and Heiner Bubb. 2005. Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. IEEE, New York, NY, USA, 56–59.

[54] Tamara von Sawitzky, Philipp Wintersberger, Andreas Riener, and Joseph L. Gabbard. 2019. Increasing Trust in Fully Automated Driving: Route Indication on an Augmented Reality Head-up Display. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (Palermo, Italy) *(PerDis '19)*. Association for Computing Machinery, New York, NY, USA, Article 6, 7 pages. https://doi.org/10.1145/3321335.3324947

[55] Marcel Walch, Stacey Li, Ilan Mandel, David Goedicke, Natalie Friedman, and Wendy Ju. 2020. Crosswalk cooperation: A phone-integrated driver-vehicle cooperation approach to predict the crossing intentions of pedestrians in automated driving. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, New York, NY, USA, 74–77.

[56] Marcel Walch, Marcel Woide, Kristin Mühl, Martin Baumann, and Michael Weber. 2019. Cooperative Overtaking: Overcoming Automated Vehicles' Obstructed Sensor Range via Driver Help. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 144–155. https://doi.org/10.1145/3342197.3344531

[57] Rachael A. Wynne, Vanessa Beanland, and Paul M. Salmon. 2019. Systematic review of driving simulator validation studies. *Safety Science* 117 (2019), 138–151. https://doi.org/10.1016/j.ssci.2019.04.004

[58] Dohyeon Yeo, Gwangbin Kim, and SeungJun Kim. 2019. MAXIM: Mixed-reality Automotive Driving XIMulation. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, IEEE, New York, NY, USA, 460–464.

[59] Dohyeon Yeo, Gwangbin Kim, and Seungjun Kim. 2020. Toward immersive self-driving simulations: Reports from a user study across six platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12.