DRIVING SIMULATION FOR INTERACTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Goedicke

August 2023

DRIVING SIMULATION FOR INTERACTION

David Goedicke, Ph.D.

Cornell University 2023

This Ph.D. thesis investigates how interactions with autonomous systems can be simulated to explore a wide range of scenarios before autonomous driving technology is deployed to the public. The advent of autonomous vehicles creates new challenges for driving simulation: user experience, non-driving related tasks, and driver-driver interactions become more relevant. Given the new challenges emerging intelligent systems bring, this dissertation develops three new simulator systems that allow interactions to unfold more freely than traditional implementations. This allows researchers to discover how and which design choices matter for seamless interaction and safe introduction to the public.

The first simulator explores how interactions between a driver/passenger and an autonomous vehicle unfold, especially in critical traffic situations; the second extended this concept to include real-world traffic. The third simulator enabled the examination of the interaction between traffic participants, specifically driver-driver interactions, and their strategies to resolve ambiguous traffic situations.

Besides enabling immersive, replicable, and reusable research, the simulators are designed to capture and reproduce rich data streams from participants' reactions. The unified view of these data streams facilitates the reconstruction of the interaction through qualitative behavioral analysis.

The thesis concludes with an outlook on how these methods and simulators, in particular the discovery-based approach, could find applications within the research fields of Human-Robot Interaction and Human-Computer Interaction.

BIOGRAPHICAL SKETCH

David grew up in Dortmund, Germany; he went to school and spent the first 20 years of his life there. In 2011, he enrolled in the BSc. Program *Creative Technology* at the University of Twente in the Netherlands. In 2014, he finished this program with honors and proceeded to do an MSc. in *Human Media Interaction* at the same university. During the program, David visited Wendy Ju's research group multiple times to develop early versions of his interaction-focused simulators that later turned into the work presented in this thesis. At the end of 2017, David finished the Master's program and started the Ph.D. at Cornell Tech a few weeks later, in 2018. During the years of his academic education, David engaged in internships at Samsung, Nissan Research (Silicon Valley), and the Toyota Research Institute (TRI).

Within time.

ACKNOWLEDGEMENTS

There is a long list of people that I would like to acknowledge as they have helped me through my upbringing and my academic education. Without the support of these people, none of this work would have been possible.

Thank you to:

My family Specifically, my parents **Monika and Michael**, my sister **Ina (and Alex)**, and **Ulla** for their continued support over the many years to help me become a person. I am deeply grateful for all the sacrifices you have made to help and support me. Your emotional and intellectual support has been invaluable to me, and I am so fortunate to have you in my life. Thank you for always being there for me.

My early teachers Who did not give up on me and continued to help me find my strength. *Johannes Blume, Dagmar Schneider, Gerhard Vögele, Karin Weber, Beata Wroblevskey*.

Educators in Twente Who gave me the opportunity to explore many different research directions, help me build local and international connections, and helped me get an early grasp on how research work gets done.*Edwin Dertien, Vanessa Evers, Jamy Li, Angelika Mader, Dennis Reidsma*.

Friends in Twente That were willing to hang out into the late hours to keep exploring and building. Our approach to teamwork and collaboration remains unmatched. *Sevim Aktas, Thomas Beelen, Mirjam van Dijk, Emiel Harmsen, Miriam Iliohan, Jan Kolkmeier, Jorik Ordelmans, Verena Zink*.

Colleagues and Friends in Stanford That supported me in this new environment, gave me a home for my visits and helped me grow my research skills. *Becky Currano, Mishel Johns, Nikolas Martelaro, Dave Miller, Srinath Sibi, David Sirkin, Hamish Tennent, Erik Vinkhuyzen*.

Colleagues and Friends in NYC Cornell Tech, and especially the Roosevelt Island campus, was still very new in 2018 when Wendy Ju started her research lab. Over the years, the lab has grown visitors passing through. I thankful for all of the people who have made this stay in New York a great experience, who helped me see the city, and who supported me in my work, and research - I would not have been able to do any of it without that support. Thank you to everyone. **In particular, I want to thank the people who held down the fort with me as the pandemic rolled over NYC.** *Alenxandra Bremers, Frank Bu, Andrea Cuadra, Natalie Friedman, Ji-Hyun Jeong, Amrit Kwatra, Rei Lee, Sam Lee, Bil Leon, Ilan Mandel, Saki Suzuki, Madiha Zahrah, J.D. Zamfirescu-Pereira, Yiran Zhao*.

Outside Collaborators During the years of my Ph.D. I was fortunate enough to have engaged in multiple collaborations. Specific thanks go to *Hiroshi Yasuda* and the folks at TRI, for having me join as an intern, during the pandemic and their support for my research through the years. Further, I need to thank the team at the Technion in Israel for their trust in willingness to use the simulators that we developed *Avi Parush, Navit Klein, Talia Wise, Carmel Zolkov*.

My Committee *Serge Belongie, Wendy Ju, Malte Jung* For their continued support to let me explore different research directions and for giving me room to grow academically.

Wendy Ju , who gave me the opportunity to start a research lab with her in 2018. Your support and guidance have been invaluable throughout the years, helping me overcome challenges in NYC. Thank you for letting me explore "different" paths and still giving me insightful advice on my research ideas. And, of course, it should be mentioned that I am very grateful for the work in securing funding for our research.

I feel anxious that there are people that I have accidentally overlooked, I apologize if that is the case.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Automotive designers and human factors engineers use driving simulation environments to test the usability, safety, and driver performance impacts of interfaces and interactions. Driving simulators allows researchers to play through scenarios to explore interactions, particularly for new technologies, designs, and critical situations. By constructing test scenarios, researchers can use driving simulators to explore the interaction between people, technology, and the environment. These systems can act as a testbed for designers, allowing them to develop and prototype for specific situations rapidly and to see how interactions between people and technology change over time.

With recent developments in vehicular technologies, systems that used to be manually controlled have become more self-directed, the likes of which we have not yet seen in our everyday lives or in a car. This change leads to new questions for automotive designers and the simulators they use. Questions of how control is exerted over a vehicle change into questions of how interaction changes the passenger/driver experience of using autonomous features of a vehicle.

In this thesis, I explore how researchers can use simulation to observe and discover these interactions. How can we construct environments and scenarios to let interactions play out and to understand better how these factors influence the development of trust and teamwork? This can be between multiple people or between people and technology. More generally, this approach of simulating rich interactive environments can be used to address a wealth of (interaction) design and behavior questions in the context of Human-Robot Interaction (HRI) and Human-Computer Interaction (HCI). While this thesis pivots around the interaction with (partially) Autonomous Vehicles (AVs), each simulator is also contextualized as a tool for broader research questions.

In particular, I have looked at how researchers can discover new interactions as they unfold with AV technologies before these are deployed in vehicles in the real world. The different systems each look at a different orbit of interaction that extends around the research subject (as driver /passenger). These "orbits of interaction" refer to the different levels or contexts at which interactions with autonomous vehicles occur, ranging from individual interactions inside an autonomous vehicle to broader interactions in the traffic environment and even specific exploration of how ambiguous traffic situations get resolved. Understanding these various orbits allows researchers to analyze and comprehend the complexities of human-automation interactions and the implications for designing autonomous systems.

In this chapter, I first discuss the contemporary context for driving simulation, as larger changes in driving and technology demand new capabilities and features for our driving simulation environments. I then discuss key features of driving simulation for research, and map out existing simulators to identify the gaps between what currently exists and what the contemporary context demands that my research seeks to fill. I then discuss different *Orbits of Interaction*, which I use to structure the interaction around AVs and use to illustrate the difference between the simulators I have developed for my thesis. Finally, outline the three projects that make up the key contributions of this thesis.

1.1 Contemporary Context for Driving Simulation

The development of new simulator technologies is based on the research questions and design artifacts motivated by newly introduced technologies that control objects in our environment. The most recent developments involve features leading to an increase in autonomy. Prominent examples of such systems are the introduction of Electric Vehicles (EVs), new “personal” transportation concepts, and advanced driver assistant systems. In this context, simulators allow designers and human factor engineers to test and perfect a system’s design before it is released to the public. As AV technologies mature, new forms of interaction with these machines will emerge. These will present unique challenges for how we think about technology use, trust, safety, and privacy. In this section, we will explore these new technologies and what new interaction design challenges they bring, with a particular focus on those that the use of simulators can assess.

1.1.1 Autonomous Vehicles

The level of automation available in vehicles has increased over the last couple of years to the point where Advanced Driver Assistant Systems (ADAS) are available in many vehicles (see [168]). These systems monitor the vehicle’s surrounding and the driver’s inputs to predict future outcomes and use that information to either inform the driver or augment the vehicle’s input to avoid/correct lousy driving decisions. The deployment of this technology is a stepping stone towards systems that can act more and more independently, leading to full AVs.

AV technology has the potential to revolutionize transportation as we know it. The Society of Automotive Engineers (SAE) has formalized these steps of increasing automation into five different levels of automation. These levels range from no automation at Level - 0 to full automation at Level - 5 (see Figure 1.1). This commonly accepted framework constructs a shared understanding of technologies’ capabilities and limitations for AVs.

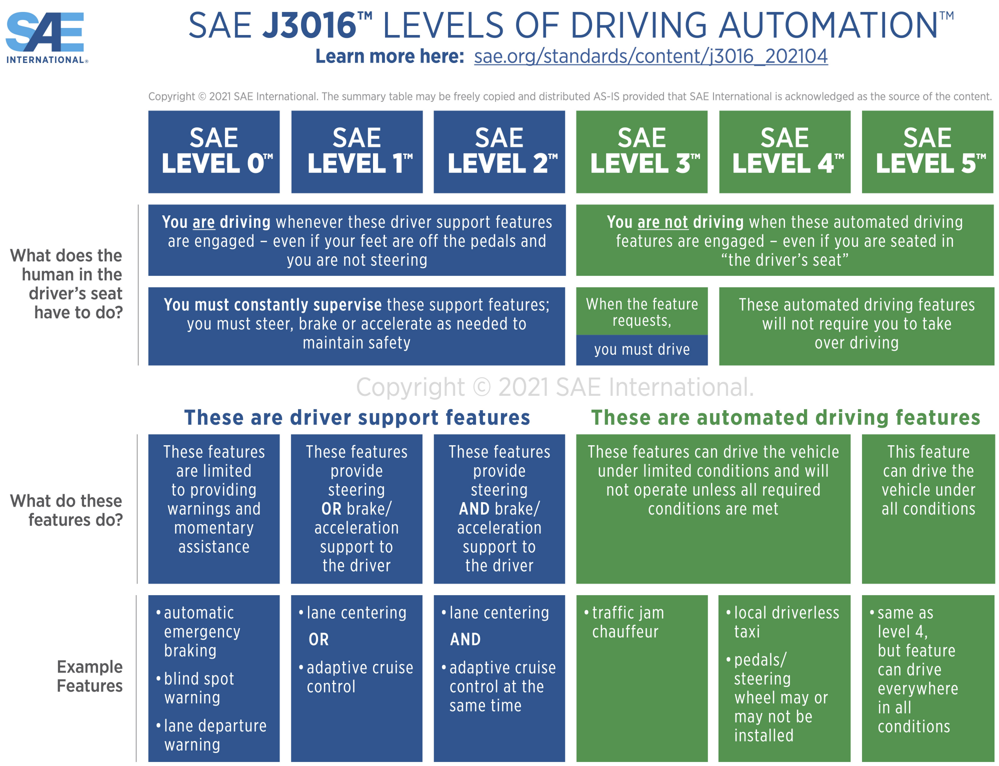


Figure 1.1: The five levels of automation with the different features and examples. *Credit: SAE International*

Levels of Automation

The automation levels are also important for work on simulators. They dictate the capability of what the simulator needs to be able to replicate. More so, it guides what kinds of scenarios the simulator needs to immerse the participant in. The requirements for the simulators can be loosely clustered into three different groups shown below:

Levels 1 & 2 For these lower levels of automation, the simulation requirements are similar to what traditional driving simulators provide. The simulation needs to be able to provide programmatic access to the participant input to augment the driving controls to prototype the behaviors of Level 1 & 2 “support” systems. Hence, “augment” the driving task but do not take control.

Levels 3 & 4 In these levels, the AV can drive parts without the intervention of a (supervising) driver. This also implies that control needs to be taken and returned to the driver potentially multiple times during a single trip. These control transitions are critical and rely on a seamless interface. To design these interactions for a trustful and safe transition of control, simulators require the ability to drive autonomously and monitor the drive and driver.

Level 5 This level of automation challenges the concept of what a car is; without the requirement to have standard control interfaces, the vehicle transportation space can change significantly. It could be used for other activities and using more communal seating arrangements. This is a new challenge for simulators, as many are built around preconceived notions about what a vehicle’s interior looks like and how a vehicle is used. Simulating prototypes and interaction design concepts for this level requires new tools to stage scenarios incorporating Level 5 AVs. Prototyping and interaction design for this level needs completely new methods and simulation to address this level of automation.

1.1.2 Non-Driving Related Tasks

Non-driving related tasks (NDRTs) refer to activities that drivers engage in while operating a vehicle but that are unrelated to driving. These tasks can include using mobile devices, interacting with in-vehicle infotainment systems,fand or engaging in various forms of multitasking. Understanding the impact of these tasks on driving performance is crucial for improving road safety and designing effective user interfaces.

As vehicles become more autonomous, the role of drivers is expected to change as well. With higher levels of automation, drivers may have more time and opportunities to engage in NDRTs. However, it is vital to study the effects of these tasks on drivers' attention, situational awareness, and response capabilities, especially for levels of automation (Levels 1 through 4). To help guide the development of the interactions and interruptions of AVs, researchers have identified potential distractions and designed appropriate guidelines for interfaces to ensure that NDRTs do not compromise safety.

Furthermore, as vehicles become more autonomous, the nature and types of NDRT may evolve. For example, drivers may shift from traditional manual tasks to more complex cognitive activities or the user of entertainment. Simulators and their methodologies must address these new challenges in understanding and managing NDRTs in the context of AV. In particular, vehicles of Level 3 and up might require participants to engage in a secondary task more intentionally to ensure they remain awake to take over control when necessary [147].

1.1.3 Other Innovations in Personal Transport

These changes in technological capabilities, especially the introduction of Level 5 automation, will likely also change how we relate to and own "vehicles". This will likely lead to disseminating more contemporary transport models such as ride-sharing [218]. The idea of ride-sharing is a transportation model where individuals share a vehicle for a specific journey arranged through mobile apps. At an urban scale, new forms of personal transport have the potential to reduce traffic congestion, improve resource use, and increase accessibility and affordability of transportation services.

As AV technology advances, ride-sharing can more easily provide on-demand transportation without human drivers. These interaction design challenges include effectively communicating the shared nature of the vehicle and facilitating seamless coordination and communication among passengers. Personal safety, trust, privacy, user control, and efficient passenger coordination are all crucial considerations for designing user-centered interfaces.

1.1.4 Implications for Driving Research

From AVs to ride-sharing, the various innovations in the transportation space bring new design and interaction challenges, especially as technology starts to act more proactively in our environment. Simulators to study some of the questions around these issues already exist. In the following section, I will show the related simulator work and highlight towards the end why new solutions are worth exploring, both for transportation work and, more broadly, HRI/HCI research.

1.2 Driving Simulation for Research

Driving simulators play a critical role in human-centered automotive research applications. They allow researchers to create safe and replicable stimuli, enabling rapid and safe empirical exploration of how people will interact with interface/behavior designs in use with transportation technology.

In this section, I will outline key features that driving simulators need to have to support driving research and discuss how existing simulators serve those purposes. As part of this work, I identify gaps in the landscape that my thesis work skills to address.

Driving simulators span from low-cost driving simulators, like City Car Driving [1] & Grand Theft Auto-based OpenIV [2], to high-fidelity immersion driving simulators, like the National Advanced Driving Simulator at the University of Iowa [60] or Ford's VIRTTEX simulator [12]. Slob (2008) and Serje and Acuña(2019) offer an overview of simulators [181, 186]. These more traditional driving simulators looked at high-risk scenarios and vehicle-interface evaluation [130] before full product development.

The two key dimensions which I will use to situate my driving simulation work are *Combined Immersion* and *Replicability/Reuse*. *Combined Immersion* is an essential criterion for the work done in simulator platforms to ecological validity [107, 189]. *Replicability/Reuse* is the cornerstone of replicable science [26, 55]. The dimension of *Combined Immersion* tries to create the dimension of immersion by combining different forms of simulated stimuli that influence a participant's immersion. Most prominently, we know of visual stimuli and motion stimuli. To simplify the comparison, I clustered the work into three separate classes of simulators. The different classes loosely build upon each other and describe an increasing level of realism that the system can achieve. Although immersion has been extensively studied, new challenges for interaction design and discovery-based work require more immersive experiences to elicit naturalistic responses to increasingly nuanced scenarios.

The *Replicability/Reuse* dimension discerns the ease with which other researchers can replicate and extend a system. For interaction research, tool and material access is a core requirement for reproducing research results and establishing a new fidelity and understanding status quo. By emphasizing these dimensions, my research addresses the gaps in the existing literature. It provides a verified first step

[1] `http://citycardriving.com`
[2] `http://openiv.com`

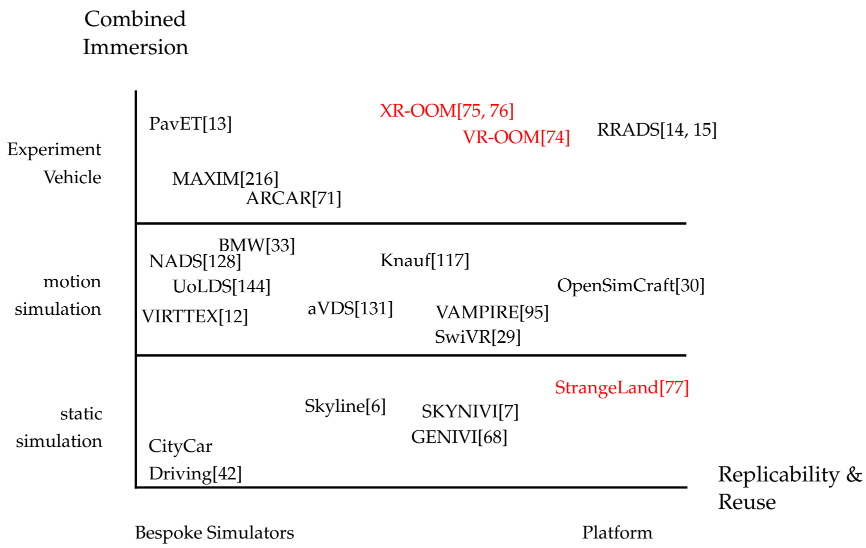


Figure 1.2: A view of the two dimensions *Combined Immersion* and *Replicability & Reuse* and how they intersection in this framework.

towards future research in behavior and interaction design for HRI and HCI and, as an instance, in AV research.

Combined Immersion

As people move through the real world, they have become attuned to drawing information out of the motion and actions of agents (other people, animals) and objects (cars, bikes) to predict intent and future actions [105]. Simulators for vehicles and robots will need to be able to reproduce these behaviors and allow for the modification of such subtle cues by designers. The simulation needs to be immersive enough to elicit naturalistic responses [16, 138] and allow for adaptable behaviors by the system to the elicited responses from the participant to let interaction fully unfold. Previous work has shown that immersion is crucial in how close to naturalistic a response can be elicited. As an example, prior work showed that perceived danger and immersion are lower and sleepiness is higher in a driving simulator than in a real car [85, 91] highlighting the need to create an as realistic as possible experience for the participant.

Figure 1.3: Simulator using an Oculus DK2 as the main Display technology. (March 19, 2015) *By Nan Palmero from San Antonio, TX, USA. Used under (CC BY 2.0)*

The *Combined Immersion* dimension references multiple stimuli that combine to create an immersive experience for the participant and collectively shape the experience of driving in a car. Visual stimuli in different levels of fidelity are always part of a driving simulator; sometimes, there is a simple computer screen, projectors, or a VR headset. Similarly, acoustic stimuli are frequently included. A major distinguishing factor in participant's immersion and system complexity motion stimuli [95]. Replicating the inertial forces (motion stimuli) is a significant challenge for driving simulation due to the technical complexity involved in replicating the felt motion. Motion replication as the dominant dimension is motivated by psychological studies that point to the importance of inertial and vestibular cues to distance perception and steering (please see [110]).

Stationary Simulators The first category comprises stationary simulators that do not provide any motion simulation. They often use regular monitors or screens and sometimes feature a vehicle cabin 1.4. More recently deployed simulators might also feature Virtual Reality (VR) (see Figure 1.3) headsets since the technology has become commercially available. These simulators can have some simulated tactile fidelity by vibrating the vehicle cabin, participant seat, vehicle, or steering. Despite these drawbacks, these simulators can be very cost-effective due to the limited hardware requirements and often can be portable, especially when deployed with VR headsets. These simulators are particularly useful for capturing a diverse participant pool of people as they can easily be transported to different locations.

Motion Replication The second group comprises simulators that use some motion simulation (e.g., [29]). The seat or vehicle cabin can be moved and rotated to replicate the sense of acceleration for the participant. These simulators' visual fidelity ranges significantly depending on their cost and specific application. Some racing simulators [1, 37] only feature a wide computer screen setup mounted to a fast-moving base, while other of the advanced motion base simulators like NADS [60] feature high-fidelity projection screens around an entire vehicle cabin. Motion-replicating simulators can quickly become expensive, especially for simulators with more than three degrees of freedom. These high-end, high degree-of-freedom driving systems that integrate driving-like motion are costly, e.g., NADS had an initial price tag of $80MM [72]. However, they only created accelerations that are a fraction of the realistic acceleration [209].

(Extended) Vehicle The third group with the highest fidelity are simulators that utilize actual vehicles as part of their systems. These simulators often use headsets [71, 216] or the Wizard of Oz (WoZ) [139] method in combination with cars driving in traffic or testing areas to create the most realistic AV simulation experi-

Figure 1.4: Static Driving simulator with vehicle cabin and three projector screens at Cornell Tech NYC. *Credit: Wendy Ju.*

ence. In-vehicle screens and/or WoZ can be used to recreate particular scenarios, while video cameras and other tracking technologies can be used to recreate the vehicle's motion for analysis later [14, 15, 137]. In particular acoustic, tactile, and motion are perfectly replicated. At the same time, the visual fidelity depends on the deployed display technology, be that a prototype interface, mixed-reality (XR) headset screen, or display.

Immersion Requirements for Simulators As mentioned in Section 1.2 many different types of stimuli exist that, when combined, create an immersive experience for the participants. While visual fidelity plays a crucial role in this work, it is essential to include additional immersive stimuli by using real-world elements to complete the immersive picture. This is to construct a convincing experience for the participant. The more closely the driving experience resembles real-world driving scenarios, the more likely we can capture *naturalistic* responses from a participant.

Replicability/Reuse

Replicability/Reuse is the second dimension to compare existing simulators. Similar to the *Combined Immersion* dimension, it is a loosely constructed scale that distinguishes the projects dependent on how much of the system's source code and blueprints is accessible and helpful for replication and extensions of the presented setup. The existing work is grouped into increasing levels of replicability and reuse. This is based on software availability, framework dependencies, and hardware agnosticism.

The replicability spectrum ranges from *Bespoke Simulators* to open *Research Platforms*, each integrating hardware and software components into a single tool. Assessing the specific reusability of simulators becomes challenging due to its dependency on technological advancements and hardware availability in the future. Thus, the clustering of existing systems emphasizes important aspects of their reproducibility.

Most replicable simulators should lean towards *Research Platforms* to facilitate a range of work in this field. For this thesis, a *Research Platform* should fulfill the following characteristics:

Accessibility Publication of hardware and software details, study protocols, and example data.

Usability Demonstration of proof of concept implementation and study.

Adaptability Implementation using modular components and popular and established frameworks (e.g., Unreal) [65, 196].

Extensible OpenSource & reusable, scalable.

Some of the key groupings for these simulators are as follows:

Bespoke Simulators are closed sources and often rely on specific hardware and software. Many of these systems exist only once and are purpose-built for certain tasks. [12, 13, 33, 42, 71, 216]

Partially open-source Many research simulators fall into this category, where parts of the simulating systems are open-source. Still, they are only designed to operate on specific hardware or subsystems. The hardware dependency, in particular, makes replication difficult. [6, 75, 76, 131, 220]

Open-Source and Hardware Agnostic Introducing common hardware components (e.g., standard displays, VR headsets) and common software interfaces to address this equipment (e.g., Steam VR, Unity, Unreal Engine) has made replication and reuse much more accessible. Not only can different types of hardware with the same software, but editing and adjusting the virtual environment is also more accessible. The major drawback of these systems is that the software is not designed to be reused for other tasks, i.e., bespoke in their implementation. [7, 29, 68, 74, 95, 117, 142]

Research Platforms These are open-source driving or traffic simulators that use standard hardware components and interfaces and are developed with a modular approach allowing for a wide of reuse for the system and its components. [14, 16, 30, 77]

In the following sections, I will describe the existing simulators mapped onto

1.2 to discuss the replicability of the different solutions. It is essential to emphasize the significance of replication, especially considering its historical challenges, as we as a community are concerned about a replication crisis [27, 49] and issues around long-term research artifact preservation [78]. As an academic community, we should strive to build systems that can be replicated as it allows for the expansion of this kind of research and promotes inclusivity in the field.

Replicability of Static Simulators

Static simulators can be some of the most replicable simulators. Some simple simulators only require a regular computer and screen. More recent implementations might use commercial VR headsets for participants to experience an immersive VR environment of the drive [17]. It is also possible to extend the virtual world experience by also placing and calibrating a physical steering wheel and foot pedals in front of a participant while they view the virtual world as it passes by.

Head-mounted displays allow changing interface elements like screens and actuators in virtual rather than physical reality, enabling a broader range of experimentation. VR tools such as the Meta Quest2 Meta Quest Pro, Pico 4, HTC Vive (Pro) LEAP Motion, and Unity and Unreal game engines are low-cost and available for a few hundred USD/EUR. However, dynamic motion (forces felt on the body) is often absent from research settings using this method. Therefore, this method has similar limitations to a mid-range driving simulator.

These VR-based driving simulations have recently emerged as an inexpensive way to have participants experience high immersions for relatively low cost; the field of view offered by VR headsets enables researchers to run urban driving scenarios, for example, which once were only possible for researchers with 270 degrees or greater field of view displays setups [121, 206].

Due to their relatively interchangeable hardware requirements, these are some of the most replicable systems. Rudimentary PC-based software simulators and screen-based simulators like City Car Sim [42] are inherently hardware-agnostic.

The primary way these simulators vary amongst the replicability dimension is in their implementation and use of common platforms and libraries. Simulators that use standard rendering engines like Unity [68, 196] or Unreal [48, 65] encourage extension and replication since these are popular tools in the gaming industry. They will likely be around to justify development efforts on simulators using these engines.

Replicability of Motion Simulators

These higher-fidelity simulators are often difficult to replicate, requiring specialized, potentially expensive hardware and software. Online communities have released Open-Source implementations of racing simulators like [30, 117]. These designs could form a foundation for a unified research platform that facilitates replicability. Future publications should broaden the perspective on their implementation of the use of different hardware components.

Using commercially available products is an alternative form of deploying a motion base simulator. Companies like [131] provide simulation systems with some motion capability but are often limited to one to three degrees of limited motion.

That said, many, if not most, higher-order simulators are bespoke implementations, like, e.g., Toyota's high-performance simulator [156] or Ford's VIRTTEX Simulator [82]. The size of these systems implies the need for special facilities (like power and space), making their more general distribution difficult. This makes these types of simulators expensive [72] and challenging to replicate. A scientific exploration of process that invite many researchers and perspectives.

For research purposes, the level of fidelity in visual scene construction or physical motion does not always need to be higher to be better; what is desirable may vary depending on the experiments the simulator is supporting. For example, a simulator used for human-factors experiments with people that have medical impairments (such as [89, 164]) might have relatively low physical motion fidelity but need to very accurately capture the reaction time of the participant to a visual stimulus. In contrast, a system to test the motor responses of older drivers might need better control and dynamics models to simulate realistic responses properly. For human-machine interaction studies, the point of driving simulators is to help designers understand users' behavioral responses to displays or notification systems in the car and events in or outside the vehicle [4].

Replicability of On-road Driving Simulation

Visual-based simulators become less effective for studies and design scenarios where drivers are only sometimes engaged in the driving task, like [147]. The level of immersion of a person staring at a tablet (in the context of a driver monitoring task) in a darkened and visual-only simulator is unlikely to be the same as that of a person engaged in a driving task. Addressing distraction and questions of attention management become more critical with the introduction of autonomous features (see Levels 3&4 in Section 1.1.1); this makes on-road simulation more critical. Using an actual car on real roads addresses the physical, bodily, environmental, and social reality that is the basis of a realistic experience [102]. It increases the transfer, fidelity, immersion, and presence of the simulated experience in a way that is difficult to replicate, even in high-end simulators. Motion sickness is also less of an issue as there is less incongruence between the perceived environment and the kinesthetically sensed environment.

VW researchers pioneered on-road autonomous driving simulation with the Wizard on Wheels protocol, using a specially reconfigured vehicle that featured a hidden second driver in the load compartment who could take over parts of the driving task, varying the degree of automation from manual control to complete automatic control [177]. More recently, Stanford researchers developed a more straightforward protocol that put a partition between a driving wizard and study participants who were given a fake steering wheel which they could use to "take over" automation [14, 208]. These vehicles are instrumented to capture the study participant and the context for each drive [46, 122] since there is an inherent variability to any study on the public roadway. The Stanford platform also permits remote Wizard of Oz interaction between drivers and remote wizards [137].

Replication of these methods depends on the research goal and what kinds of designs are supposed to be evaluated. Solutions like the approach by VW [177] require significant technical expertise and material to deploy safely. These are prerequisites most universities cannot fulfill. On the other hand, deploying a WoZ system like *Daze* [139] deployed by Stanford does not require special equipment and achieves the simulated driving experience through clever placement of a visual barrier between the driver and participant.

1.2.1 Measuring Driving Interaction

Most immersive driving simulators aim to elicit naturalistic responses from the participant. In this context, an essential functionality of a simulated then is to be able to record these reactions and interactions as they unfold. The simulator's output, i.e., the raw data streams, must provide enough granularity to reconstruct what and why a scenario unfolded. Looking at traditional simulator work, many measures already exist (e.g., *Steering Wheel Reversals*, or *Time to Line Crossing* [101]). These measures have been traditionally designed to provide data for questions of control and distraction, often described as *driving performance measures*. They come from a time when drivers were the sole controlling entity in a car or truck.

Early work in automotive gestural interfaces [5] and speech interactions [70, 83, 125, 150], for example, make use of the driving simulator and focus on driving as a primary task that designed interactions should not interfere with.

Driving Performance covers a whole host of measures that attempt to assess a driver's performance in a vehicle. Many of these measures are standardized and published in SAE J2944TM [101] after being refined and validated in previous academic research. A portion of these measures looks at control problems. With this control approach, these measures fall close to traditional driving research, where distinct conditions lead to specific, often predefined outcomes.

Some examples of *Driving Performance* measures are *headway distance* [101, p.50], *response time* [101, pp.38-43], and *steering entropy* [101, p.137]. These key performance indicators are used to compare different study conditions, to evaluate vehicle designs and potential traffic rule changes. These measures are particularly useful when research questions have been clearly defined. So much such that regulations on the control of vehicles make use of these definitions [101, p.8]. Such that the impact of new vehicle designs (e.g., media center) can be evaluated and compared using these predefined metrics and deemed safe.

However, as we move towards modern vehicle technologies, these traditional methods and approaches become less relevant; some of the measures presented in this J2944TM remain just as relevant when looking at the interaction between people and AVs and should always be included when possible. In particular, measures that assess attention, NDRT, and general passenger experience are still relevant.

Driving Psychology

In the field of driving psychology, driving behavior models have been examined to understand the mechanism of accidents caused by human errors in particular [113, 146]. According to these models, driving can be clustered into three hierarchical levels: strategic, tactical, and operational. These levels correspond to their information processing demands, similar to Rassmussen's skill-rule-knowledge (SRK) behavior model [171]. The tasks for each level are described as trip planning, maneuvers to handle prevailing circumstances, and low-level lateral and longitudinal controls, respectively.

However, when examining how interactions between people and technology unfold, these existing and predefined measures do not necessarily cover the nuance required to reconstruct the causal signaling that leads to the interaction (e.g., when negotiating ambiguous traffic situations). Simulators need to provide data streams for measures that, more openly, can lead to discovery-based research such that serendipitous and unfolding interactions can be reconstructed.

Driving Simulation for Autonomy and Driver Assistance

AVs bring new concerns, requiring new models for visual display systems, control interfaces, audio alerts, and interaction [3, 58]. Many experiments for driving simulation involving automation are controlled "transition of control" studies [80, 145]. However, some, such as [151, 176], have taken a more designerly and improvisational approach to sharing or transitioning control with automation. This design- and development-oriented use of driving simulation are also gaining traction for non-automation uses; the Intel Skyline simulator, for example, is focused on making it easy to prototype interactions between the vehicle and brought-in devices such as phones and tablets [6].

1.2.2 Measuring Qualitative Experience

Qualitative measures (e.g., open-ended interviews and behavioral coding) provide a valuable means to capture naturalistic responses in research and offer distinct advantages over quantitative measures. Unlike quantitative measures, which often rely on predefined hypotheses and structured data collection, qualitative approaches embrace open-ended exploration and allow a deeper understanding of human interactions with technology and design.

Participants can naturally interact with the technology by developing unconstrained environments in sensory-rich simulators. Researchers can then observe and analyze how interactions naturally unfold without imposing predetermined constraints. This approach enables a more authentic exploration of how interactions take place in real-world contexts, offering insights into the complexities and nuances of human behavior and technology interaction.

Open, discovery-focused methods make it possible to uncover serendipitous combinations of events to test and discover new ways in which interaction unfolds. In contrast to more traditional simulators and measures, the simulators presented in this thesis utilize real-world influences on how the scenarios are constructed and how they are let to unfold. This open approach to creating scenarios with open-world influences allows researchers to explore new and unforeseen combinations of scenarios. The main drawback is that these methods' findings are often not easily generalizable.

Methods such as behavioral coding allow us to systematically analyze participants' behaviors, discern patterns, and uncover the cues they attend to and interpret. Through this approach, we gain insights into the dynamics of interactions, identifying subtle nuances and contextual factors that quantitative data alone may overlook. This more open and discovery-based research approach enables us to explore and comprehend the multi-faceted nature of human-automation interactions after specific interactions have been identified. They can sometimes be codified and turned into new quantitative metrics.

As the field moves towards interaction design questions with the introduction of Level 3 & 4 autonomous vehicles, requirements for driving simulators change. One could expect that other immersive dimensions, other than refining driving physics, will gain more attention, like immersion and scenario flexibility, become more important.

1.2.3 Opportunities for Simulators

The existing research in the driving simulator field exhibits some gaps that provide room for my contributions. Firstly, current simulators tend to be high-fidelity, complex, and expensive, relying on specialized hardware to replicate the felt motion to a limited extent. This cost and complexity make replication of work at other research institutes challenging, which restricts accessibility for researchers and limits the communities' ability to validate published results.

Secondly, the open discovery-based approach, which allows for serendipitous interactions between people and AVs to unfold, remains largely unexplored in simulator work. By integrating designed simulated scenarios with real-world traffic environments, my work captures a wide range of scenarios and unveils unforeseen combinations, leading to novel insights into technology design.

To enable this kind of open, discovery-based research, the simulators presented in this thesis all incorporate methods for capturing the richness and complexity of these interactions while driving. They are to provide the means to uncover deeper insights, understand participant perspectives, and explore the intricate interplay between human behavior and their surroundings, be it other human traffic participants or AVs.

1.3 Driving Simulation for Different Orbits of Interaction

New vehicle technologies bring new challenges for interface and interaction design. All the while, graphic and computational technologies also become more capable. New simulators can be conceived as devices that help explore possible future interactions.

With three distinct projects, I look at how interactions with and through (autonomous) vehicles can take place. Each method has its own "orbit" within which deployment is beneficial. In this document, I present them in a growing fashion. From the simulation of scenarios *"inside an AV"* to navigating through the environment *"around the AV"* to interactions with *"Other Road Users"*. The following section introduces these different *orbits of interaction* before the following chapters describe each project in more detail. Furthermore, it is described how other researchers have picked it up.

All simulators facilitate a discovery-focused research methodology, and while in this incarnation, they are considered interaction with AVs, lessons, and concepts can be carried over to other kinds of interaction research questions with automation, transport, and robotics. This prospective applications are described in the final chapter (see Chapter 6).

1.3.1 Inside an Autonomous Vehicle

After the publication of *VR-OOM* as a system [73] and its evaluation as a user simulator platform [74], work continued on development and testing to improve the system's key performance factors (e.g., tracking reliability and ease-of-use). These improvements led to new capabilities; while before, *VR-OOM* was only capable of running in large open areas like a black lake (see study in [74], the current version allows for testing on regular roads and in traffic (when a researcher drives the car).

In its various iterations, *VR-OOM* explores how interactions inside the car could occur and how this method can evaluate people's behaviors and reactions inside an AV unfold. As part of this thesis, I briefly discuss the original system, its shortcomings, and how it was improved and conclude with a description of the capabilities of the improved system.

1.3.2 Around the Autonomous Vehicle

In the next phase, the research focused on evaluating interactions between a driver outside environment. To afford a perspective on real-world objects, the simulator needs to support the pass-through of the visual elements, which requires an XR headset that can show both the real and virtual worlds simultaneously. This means the participant will see an extended or augmented visual presentation of the natural world's surroundings.

While similar to *VR-OOM*, the deployment is significantly more challenging due to higher accuracy and refresh rate requirements than in a VR application. The benefit is that it does allow for much more realistic scenarios and even allows for testing where the participants can drive. This Mixed Reality(XR) implementation of the *VR-OOM* simulated was extended to include and renamed to *XR-OOM* [76]. Early implementation tests were published under [75].

1.3.3 Other Road Users

For AVs to adapt to interact with other drivers and to adapt to local norms, it is critical to understand how those norms differ and to profile how it varies across geographical locations. While ethnographers have qualitatively described regional differences in driving style, AVs would need data-driven statistical models for AVs to recognize how local drivers are signaling through hand/body movement and motion of their vehicles.

1.4 Outline of Thesis Research Projects

For this thesis, I have developed three different driving simulators that address each of the different orbits of interaction previously outlined. The first simulator, *VR-OOM*, explores how VR interactions between a rider and an AV unfold, providing system details and implementation suggestions for replication. *VR-OOM*'s main contribution is of a technical nature providing insights into the technical challenges of deploying such a system and the required performance from tracking and VR systems.

The second, *XR-OOM*, extends the concepts from *VR-OOM* concept to include real-world traffic. The main contribution of the simulator is the constraints that I describe that emerged as the pilot study results. I also published *XR-OOM*'s software, SOP, to aid in the replication of this work.

The third simulator, *StrangeLand*, then looked at the interaction between traffic participants, specifically driver-driver interactions, and their strategies to resolve ambiguous traffic situations. *ReRun*, initially developed as an addition to *StrangeLand*, was used for post-study analysis of such studies. With both systems, we have created a research platform that can be used for a whole host of AV and HRI-related research questions. The platform was evaluated with multiple studies and has been refactored to become more stable, provide more data, and support more simultaneous clients.

To disseminate approaches to discovery-focused research methods and to encourage replication of this work, the systems are developed to use openly available software and affordable hardware component. Furthermore, the publications detail the use of the simulators and include shared resources such as links to source code and study materials and code used for analysis. (Please see Appendix A.1 for links to the source code used in this project.)

CHAPTER 2

THE EXPERIENCE INSIDE OF AN AUTONOMOUS VEHICLE

Figure 2.1: *VR-OOM* allows participants to experience the physical sensations of the real world with controlled virtual environments and events.
Photo by Arjan Reef.

The closest orbit of interaction assesses how drivers or passengers inside a (semi) autonomous vehicle (AV) experience their trip. How, in particular, critical traffic situations are being perceived, and how designed behaviors of the vehicle and its interior play a role in that perception. The Virtual Reality- On-rOad driving siMulation (*VR-OOM*)[1] system allows for prototyping for interfaces and driving experience inside of the vehicle.

In the following chapter, we will discuss the premise of *VR-OOM*, what it set out

[1]This chapter reuses material from the original CHI publication [74] and master thesis [73]. Those works were co-authored with *Jamy Li, Vanessa Evers, and Wendy Ju*, but I was the lead author of those papers and the primary researcher on that body of work.

to do, and how it was developed further after its initial publication. We demonstrate its current capabilities to simulate scenarios and to immerse the participant in a test scenario. The chapter concludes with a look at possible future work and work that has cited the original publication.

2.1 *VR-OOM*: Original Overview

The original *VR-OOM* system and research protocol enable researchers to run on-road studies with controlled events, simulate autonomous driving in a higher fidelity environment, and prototype a wide range of human-vehicle interactions and interfaces. In the paper [74], we provide a detailed description of *VR-OOM*'s system design and setup and an initial validation study of the system. Similar to the other methods discussed in this thesis, *VR-OOM* is relatively low-cost, as it uses consumer-grade entertainment and gaming hardware within a standard passenger vehicle. The system was open-sourced, including the software, 3D models, and course designs of *VR-OOM*, to lower the barrier for other automotive user interface designs to engage with experimentation and research to increase the variety, quantity, quality, and safety of the systems and interactions created.

The original implementation of *VR-OOM* had several drawbacks that impacted its usability and application range. Firstly, it had limited tracking capabilities as it relied on sensors with limited accuracy. In particular, speed was captured using OBD2, which has build inaccuracies, while the vehicle's orientation was measured using an early-generation integrated circuit IMU with high drift. This drift-prone implementation restricted experiments to empty parking lots where the driver had to compensate for the virtual vehicle's predicted path.

Secondly, the research driver had to simultaneously observe the real world to avoid obstacles while monitoring a secondary virtual screen displaying the view from the virtual vehicle. This dual tasking imposed a high cognitive load on the driver, limiting the range of scenarios that could be explored.

Moreover, the requirement of using an empty parking lot constrained the type and size of the scenarios that could be used even further.

Additionally, the tracking of the (virtual Reality) VR headset inside the vehicle was limited due to the lack of easy integration with VR software at the time. Only rotational tracking of the headset was possible, neglecting positional tracking, although this limitation was deemed acceptable as head movement relative to the vehicle was minimal. Addressing this limitation became crucial in subsequent versions.

Lastly, the graphical fidelity was restricted by hardware limitations, especially as VR-ready PCs and laptops suitable for operating inside a moving vehicle (given the limited power availability) were not widely available then.

Many of these limitations were acceptable for the original implementation of a prototype to test its use as a simulator. Furthermore, many of these limits were also due to the limits of the (affordable) technology at the time. Especial spacial tracking hardware for VR applications has seen significant advances since the introduction of the Oculus DK1, the headset that starts the current VR-hype cycle [185].

2.1.1 In-car VR as a Research Platform

CHI researchers have recently investigated the possibility of using VR in cars with head-mounted displays (HMDs). The CarVR system [93] tracks a non-virtual car's motion and renders the corresponding visual perspective of a passenger in the virtual space, which is used to play an arcade-like shooting game. The authors found that moving the game in concert with the car's motion caused less discomfort than playing the game while the vehicle was parked. McGill et al. looked more carefully at how the correspondence of motion between the visual display using HMDs and the car's motion affected motion sickness [141]. They found that motion sickness would represent an obstacle to using VR in the vehicle in real-world conditions. However, both research projects used smartphone-based VR. Honda's DreamDrive [19], demoed at the 2017 Consumer Electronics Show in Las Vegas in January, suggests that many of the issues with these research systems can be circumvented by using higher-end VR systems with higher visual refresh rates and by using the car's CAN bus data to more accurately map the virtual world movement to the vehicle's actual movement.

VR-OOM was inspired by military flight simulation. Bachelder et al.'s Fused Reality system [13] enables pilots wearing VR headsets to fly real planes in real skies while experiencing simulated situations. This system is used to help pilots practice take-offs and landings 30,000 feet in the air, where there is no threat of ground collision. It can also be used to simulate mid-air refueling or formation flying without the danger of mid-air collision. Fused Reality provides a higher-fidelity simulation environment than ground-based simulators because the aircraft is real and in motion; only what the pilot sees is virtual [32]. To leverage the benefits of Fused Reality in the in-car VR research space, we created *VR-OOM*. This novel low-cost virtual reality system operates in a moving car where the car's physical motion is mapped to the virtual road environment.

VR-OOM introduces a novel in-vehicle driving simulation system that takes advantage of low-cost virtual reality technology breakthroughs to create more im-

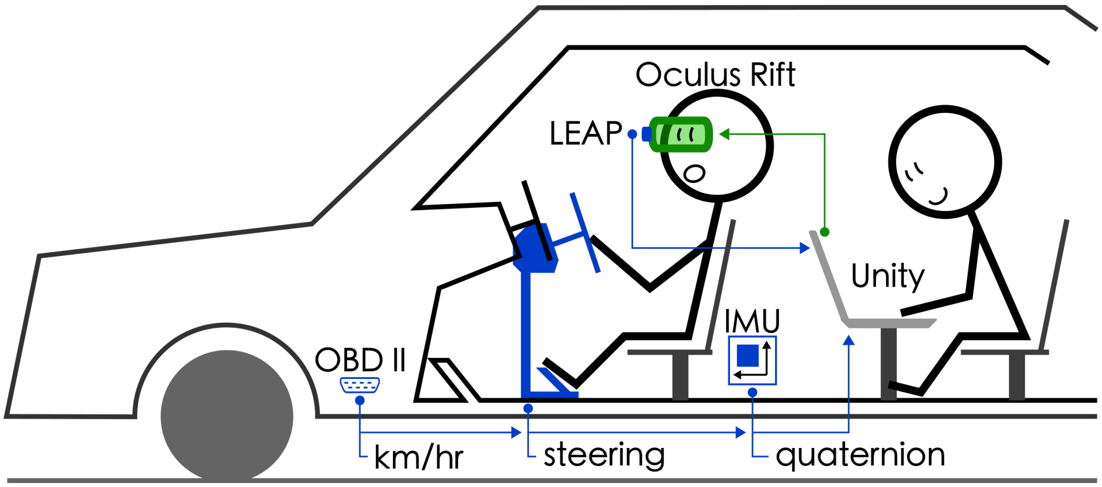


Figure 2.2: *VR-OOM* System Diagram (excluding the wizard driver)

mersive experiences than a traditional lab-based driving simulator, allowing for greater flexibility in the test environment than everyday on-road driving. This technology enables us to extend on-road driving simulation by injecting virtual objects, interfaces, and environments into the driving context, fusing the physical reality of the car with the simulated scenarios we have created.

The *VR-OOM* system and research protocol enable researchers to run on-road studies with controlled events, simulate autonomous driving in a higher fidelity environment, and prototype a wide range of human-vehicle interactions and interfaces. We provide a detailed description of *VR-OOM*'s system design and setup and an initial validation study of the system. *VR-OOM* is relatively low-cost, as it uses consumer-grade entertainment and gaming hardware within a standard passenger vehicle. We are open-sourcing the software, 3D models, and course designs of *VR-OOM*. We hope that lowering the barriers to automotive user interface design and experimentation will increase the variety, quantity, quality, and safety of the systems and interactions created.

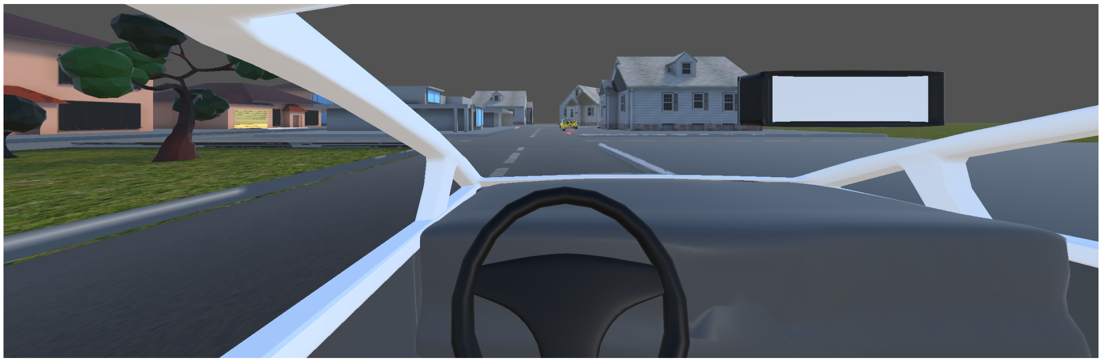


Figure 2.3: View from the participant's point of view.

2.1.2 Critique of Initial Publication

Critiques of the system mainly concerned the quality of the graphics in the VR rendering. At least three participants commented on this. The user experience of *VR-OOM* can be enhanced by providing high-fidelity graphical rendering. However, we were able to get natural responses to traffic events from participants with the current low-fidelity graphics (see Figure 2.3). Therefore, to assess responses, the graphics quality may not be essential. Fortunately, improving upon this aspect of *VR-OOM* is relatively straightforward. It involves the digital design of a consistent and complete environment filled with assets like houses, cars, people, benches, trees, and other objects. Additionally, aspects such as light rendering in the virtual world could be improved by fine-tuning the build in Unity.

Practical Improvements

Other critiques of *VR-OOM* concerned the physical components of the platform. One participant mentioned, "*I am missing the pedals. After the last round, the driver was parking very close to the other car, and I noticed I was pressing my foot down. Just to brake. The reflex of pressing my foot down.*" People who drive regularly may incline to press the foot pedals or switch on a light. The current study focused on participants' tendency to grab the steering wheel. From this comment, other reflexive behaviors can likely be similarly investigated.

In the current system implementation, the VR-Camera is fixed relative to the car after the calibration, but the rotation is decoupled between the headset and the car. This means that the vehicle and the headset determine their orientation independently. The noise in the sensors used to determine orientation cause both components to drift apart slowly over time. This is why it was necessary to recalibrate the participants viewing direction after every condition in the proof of concept implementation.

This issue could be circumvented by using higher-quality IMUs or by implementing a tracking solution to determine the VR headset's position and rotation relative to the car. Consumer grade room tracking solutions cannot operate in a moving reference frame, so the system needs to extend open source solutions like OSVR[2] or use other tracking methods like marker tracking. The addition of positional tracking would allow the participant to move their head around in the car interior more freely.

Latency is another aspect that affects the quality of a VR experience. Suppose the latency between head motion and the displayed image becomes too great ($> 75ms$ [207]). In that case, the participant's perception and motor control will be affected, influencing how naturally they can act in the virtual environment. How the delay between the tracked car motion affects well-being and immersion is unclear. The frame rate of the VR operating system also affects latency. While 90 frames per second (f/s) are typically recommended for VR [44], this project ran at about $60 f/s$. This was due to the computational overhead of the Wizard View. Future implementations will need to address this either through native plug-ins

[2]Open Source Virtual Reality http://www.osvr.org/

for video capture and better timing within the frame [3] or simply by using more powerful computing hardware.

Participants' qualitative experiences and the experimenter's observations indicate the applicability of a low-cost Fused Reality car simulator such as *VR-OOM* to assess people's genuine responses to driving situations. This first pilot study also offered directions for technical improvements to increase the immersion and practicability of the system. Moreover, this proof of concept validation provides insights for applying *VR-OOM* for on-road testing and development of AVs.

2.2 Improvements for Scenario Flexibility

Initial development on *VR-OOM* focused on addressing previously established drawbacks. The improvements include several changes.

Firstly, better hardware was used to improve performance. The IMU, the core tracking sensor of the system, was replaced by a more performant version. The new hardware could be calibrated, specifically the MTI-300 by Xsens (now Movella) with Automatic Heading Correction. This significantly reduces the drift of the sensor, keeping the headset orientation consistent vehicle's reference frames. Additionally, hardware improvements allowed for higher visual quality through higher-quality rendering and better (higher density) displays built into the headset.

Secondly, a simplified model of Roosevelt Island, the area, was modeled as a geometric ground truth in which the sim was deployed. This model serves as a validation tool for tracking and system performance.

[3] https://medium.com/google-developers/real-time-image-capture-in-unity-458de1364a4c

Thirdly, the OBD2 sensor for speed measurement was replaced with data provided on the vehicles' CAN bus from the internal odometer. This change improves the accuracy and responsiveness of the speed reading.

Lastly, a real-time manual adjustment mechanism was added to allow researchers to correct tracking errors and control the participant's experience while driving through the simulated world. Ultimately, these improvements enable the use of vehicles on regular public roads instead of being limited to specific empty areas or parking lots.

2.2.1 Roosevelt Island's First Digital Twin

For *VR-OOM*, the use of a digital twin is beneficial for a variety of reasons. Firstly, it allows for a virtual representation of the physical bounds of a test track (in our case Roosevelt Island). Changes and modifications to the study track can be tested before implementing them in the real world. The original digital twin used in *VR-OOM* was modeled after satellite imagery. It was later extended many times and now finds its way into various other projects at Cornell Tech.

2.3 Broader Perception of the Work

Since the original publication in 2018 [73, 74], the community has published other papers and systems that use and improve the original publication.

In particular, a publication by Yavo-Ayalon et al. [215] used the concepts from *VR-OOM* and the digital twin to create a visually augmented bus trip on the Roosevelt Island bus. The system was used to see if "Situated and shared experiences can motivate community members to plan shared action, promoting community engagement." [215]. See Figure 2.6.

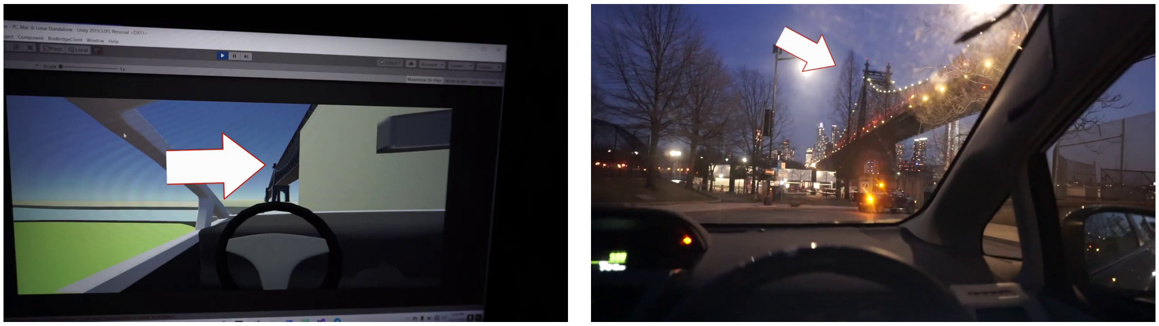


Figure 2.4: This figure shows two images from a simulation run from *VR-OOM*. The picture on the left shows the view outside of the *VR-OOM* system onto the Roosevelt Island digital twin. The picture on the right shows the same view of the real world. The references have been generated from a video showing the system working. See `https://youtu.be/nb574wfbyos`

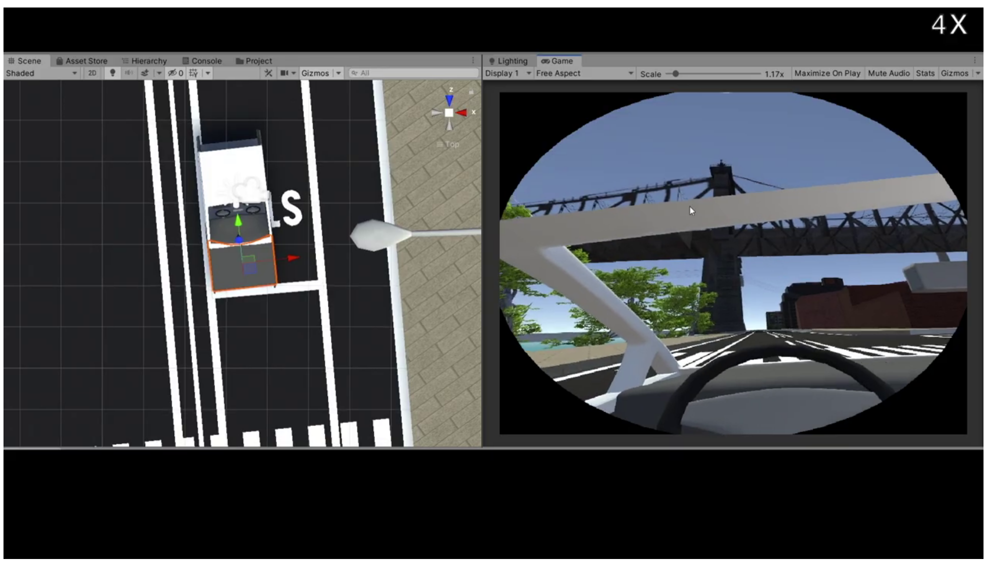


Figure 2.5: Screenshot from the unity editor showing both a top-down view showing the virtual Prius on the road and the view onto the road from the participant's point of view.

In some of the responses from the Autonomous Vehicle (AV) research community, some implementations used 360°video [69], and others developed similar systems on other VR hardware [216]. Some of the cited work also stated exploring HRI-related topics highlighting the applicability of these methods outside of the AV context. Simulating robots and their interactions with people could be an easy early discovery step for interaction design.

Figure 2.6: Yayo-Ayalon et al.'s CXR system [215] built off of VR-OOM's Roosevelt Island digital twin.

By implementing open-world rules and object behaviors into the VR world, we can let scenarios unfold naturalistically, extract behaviors from participants' interaction with that virtual world, and use it as an essential source of information for good interaction design.

2.4 Conclusion

The work on the *VR-OOM* system since its original publication [73, 74] was to extend its operational capabilities. In this chapter, we discussed both the original implementation's drawbacks and the technical solutions we found to address these challenges.

The main contributions for *VR-OOM* are both technical and methodological. First, *VR-OOM* makes evident how vita and difficult solving the stacked tracking problem is for a successful deployment. The addition of higher quality sensors that included automatic heading correction lower sensor noise and allow for deployment on real roads. However, this system still required adjustment controls for the lane position to deploy *VR-OOM* successfully. It became clear that for a successful deployment of such a system, and especially for deployment with mixed-reality (XR) technology sensors working in an absolute reference frame would be required, as the relative-tracking method was simply too noisy. Early and separate experiments establishing the tracking solutions' accuracies is a vital step in deploying such a *VR-OOM* system. Second, the methodological advancement stems from early tests with the system. As the *VR-OOM* system was tested on real roads, it became apparent a digital reference was needed to further build out the capabilities of the system and design simulation scenarios. This was achieved by modeling a digital twin of the testing environment. The digital twin of Roosevelt Island was subsequently used in *VR-OOM* and *XR-OOM* studies. It allowed for the exploration of how open-ended, naturalistic testing in real traffic can be combined with VR-based simulators. This digital twinning concept that was rigorously evaluated in this project but an essential element of the *XR-OOM* project discussed in the next chapter (see Chapter 3).

CHAPTER 3

DRIVING THROUGH REAL WORLD

Letting in more of the real world.

Figure 3.1: *XR-OOM* enables experimental studies where participants encounter virtual objects (cones) while driving a real car.

For the second orbit of interaction, our goal is to let more real-world elements influence and construct the simulated scenario to discover new and unforeseen interactions. To this end, this chapter discusses the *miXed Reality On-rOad siMulator* (*XR-OOM*)[1], a simulator that extends *VR-OOM* by bringing in substantial real-world elements (see Figure 3.1. This system opens the orbits of interaction to include the real-world traffic environment in the studies scenarios. It introduces realistic road topology, traffic scenarios, and other road users that are essential to experience the ride in an autonomous vehicle (AV). Depending on how it is deployed, *XR-OOM* can test a wide range of AV-focused and other scenarios.

This approach allows for the open-ended reconstruction of scenarios as they unfold in interaction with the natural world and includes accommodation for both qualitative behavioral analysis and quantitative methods.

[1]This chapter reuses material from the original publications [75, 76]. Those works were co-authored with *Alexandra W.D. Bremers, Sam Lee, Fanjun Bu, Hiroshi Yasuda, and Wendy Ju*, but I was the lead author of those papers and the primary researcher on that body of work.

3.1 *XR-OOM* Introduction

Automotive designers and human factors engineers use driving simulation environments to test the usability, safety, and driver performance impacts of interfaces and behaviors of the working systems with people. Driving simulators can help recreate scenarios that may rarely occur in the world or would pose an unreasonable risk to study participants on the stage. By controlling experimental procedures and scenarios, driving simulators can help recreate the same factors and events within and between experimental conditions. In addition, driving simulators can act as a test bed for designers, allowing them to rapidly develop and prototype for specific situations and environments and see how people might behave in those contexts.

It is common to use a mix of high and low-fidelity test environments throughout the development of vehicular systems [57]. Current-day fixed-base simulators in laboratories usually feature screen-based instrument panels and center consoles to enable the rapid development of user interfaces. They are authored in graphical simulation environments (such as Unity or Unreal) to make it easy to create contexts and scenarios. The level of fidelity of driving simulators has been found to affect drivers, such as the presence of a realistic cabin leading to conservative driving styles and reduced motion sickness [22]. Test-track environments are more challenging to develop interfaces and scenarios than simulators but feature more realistic driving experiences such as vibration and physical motion. More naturalistic tracks and the real road studies' pass-through of the real-world environment also help clarify the impact of external factors that designers might otherwise be blindsided by.

This contributes a critical step towards safer, less expensive development environments for designing and testing future vehicle interfaces. Using a mixed-reality (XR) headset in a real-world test vehicle, simulated events and interfaces can be tried while drivers operate and experience being in an actual car. To reduce the efforts of replicating the system and creating a realistic driving experience, part of this thesis includes sharing the specification and validation of our system to make the approach accessible to the broader automotive research community.

We seek to combine the strength of immersive simulation environments and real-world testing environments through the use of XR driving simulation. In this paper, we validate the safety and viability of this combination towards the goal of using mixed reality to simulate driving scenarios with advanced driver assistant systems. In Section 3.2, we discuss the design requirements to make in-vehicle XR systems useful for driving simulation experiments in research and design. Next, in Section 3.3, we describe the mixed reality system we built to operate driving simulation experiments in vehicles.

Section 3.4.3 outlines safe and effective operational procedures and protocols we used to validate the systems. We validated the mixed reality system for basic driver cockpit and low-speed driving tasks, comparing the use of the system with non-headset and headset-only driving conditions to ensure that participants behave and perform similarly using this system as they would otherwise.

3.2 Specific Requirements for Mixed Reality

3.2.1 Visual Resolution

The visual resolution influences the level of immersion of the participant and their ability to acquire situational awareness. Technically, this is bound by the screen's resolution and the cameras that capture the environment.

One crucial functional requirement is the participant's ability to see and comprehend visual information, particularly reading text. E.g., on navigation displays or instrument panels in the vehicle. Text readability is challenging in virtual reality because it requires very high display resolution [84, 116]. Prior research by Dingler et al. indicates that the resolution needed for text readability in VR is a function of the text size and render distance [45]. Curevo et al. found that "life-like" virtual reality requires a resolution of 60 pixels per degree for 20/20 vision [35].

3.2.2 Field-of-View

Whereas visual resolution concerns the number of pixels per angular degree-of-view, the field-of-view (FoV) requirement is a function of peripheral perceptions. Van Erp and Padmos argue that FoV is critical to lateral control tasks, leading to a more substantial impact on driving performance than visual resolution or latency [204]. Driving performance was better in drivers with a 100° FoV than those with a 50° FoV. Hu et al. note that the human eye has a 210°×150°(i.e., diagonal 200° [35]) FoV. Most commercial VR headsets only achieve a maximum 150° of FoV[2]. The most critical zone, however, is the 60°around the center of the view [97]. Similar to the visual resolution to resolve details, the FoV requirement depends to a certain extent on the experimental task the simulator supports. Any tests which draw on the participant's situational awareness or response to out-of-vehicle cues or events would need a broad FoV. Tasks that take place in a slow traffic environment or center on in-vehicle interaction may use just a subset of a typical human FoV.

[2]The Headset with the highest FoV that we could find is the Vision-8k-x with an FoV of 200° `https://pimax.com/product/vision-8k-x/` Retrieved January 8th, 2022

3.2.3 Pass-through Latency

Pass-through latency describes the combined latency of capturing the real world with a camera and projecting it onto the screen with all processing steps in the middle. Research on the combined latency effects on driving performance in simulated teleoperation of vehicles indicates a delay of 300 ms degrades driving performance [160]. Related research looking at network-induced latency jitter shows that jitter is a stronger predictor of negative remote driving performance, but also found a performance threshold at the 350 ms range [133]. Studies on networked multiplayer video games, particularly racing games, set the acceptable latency range at 50 ms [28, 166]. With longer delays between participants' control and visual response, the perception and performance of game players start to degrade noticeably.

However, mixed and Virtual Reality applications generally must meet stricter latency requirements than in remote operation and multi-player gaming [35]. Mania et al. identified the Just Noticeable Difference point to be at about 15 ms, with up to 20 ms being referenced as the maximum amount of delay before performance and immersions start to be impacted [135]. Hu et al. indicate that the VR interaction latency should be less than 20 ms to avoid motion sickness and discomfort [97]. Hence, we believe that systems that participants can operate in without feeling motion sickness or discomfort should be reasonably good at supporting driving operations and on-road coordination with other drivers.

3.3 System Design

Our system, *XR-OOM*, establishes the possibility of using XR for experimental driving simulation research. (See Figure 3.2.) It is built using the Varjo XR-1 XR-headset and Unity simulation environment, which previous researchers demonstrated could be usable for driving simulation in a moving vehicle [71]; *XR-OOM* integrates and validates these into a usable driving simulation system.

This section motivates the setup based on prior immersive XR and driving simulation systems design. These requirements address the usability of such a system and the participant's safety, well-being, immersion, and comfort. This summary of requirements should help replication attempts to evaluate other potential software and hardware to be used in such systems as technology progresses and other solutions become viable.

The current implementation is then but a snapshot of current technological limits, and we anticipate that the viability of such a system will increase as technologies improve. The validation of this prototype can show what kind of research is possible at the moment while highlighting what specific advances might be beneficial to extend the simulator's operational range.

3.3.1 Tracking and Localization

The system must solve a nested tracking problem for the simulator to render objects correctly relative to the car and the outside world.[3]. We use the SmartTrack3 from ART [148] to track the participant's head position and a ZED 2i Stereo Depth Camera [188] mounted on the front bumper of the research vehicle for in-world location (see Figure 3.2).

[3]For example, in [148] blog post, the Hololens team notes recently being able to perform similar things for boat motion and mentions explicitly that such performance is not yet possible for rendering virtual objects in moving systems like cars.

In-car Localization

The head pose of the participant needs to be tracked relative to the vehicle cabin to accurately render the in- and out-of-vehicle scenes from the driver's perspective. This makes it so that virtual objects within the car- for example, novel instrument panel displays or in-world study questionnaires or instructions- can appear fixed even if the participant moves their head.

In-world Localization

Localizing the car's position and movement is critical to properly anchoring virtual objects into the real-world scene outside the vehicle. Combining in-car localization with vehicle localization gives us a continuous estimate of the headset's location and orientation within the vehicle and the world.

3.3.2 Additional Hardware and Software

Hardware Configuration

Details of this hardware implementation are detailed in Appendix A.2.1. System diagrams, code, and simulator data for *XR-OOM* are on GitHub (Appendix A.1). Earlier publications about this approach have not previously documented the details of the required system configuration. They are necessary for any researchers using such a system to perform research experiments with [71].

Custom Software

Built on top of the Unity game engine, We developed software to render virtual objects in the real world outside the car and to obscure those objects when the physical car interior would have occluded the physical counterparts of these virtual objects. This required synchronization between the location of the partici-

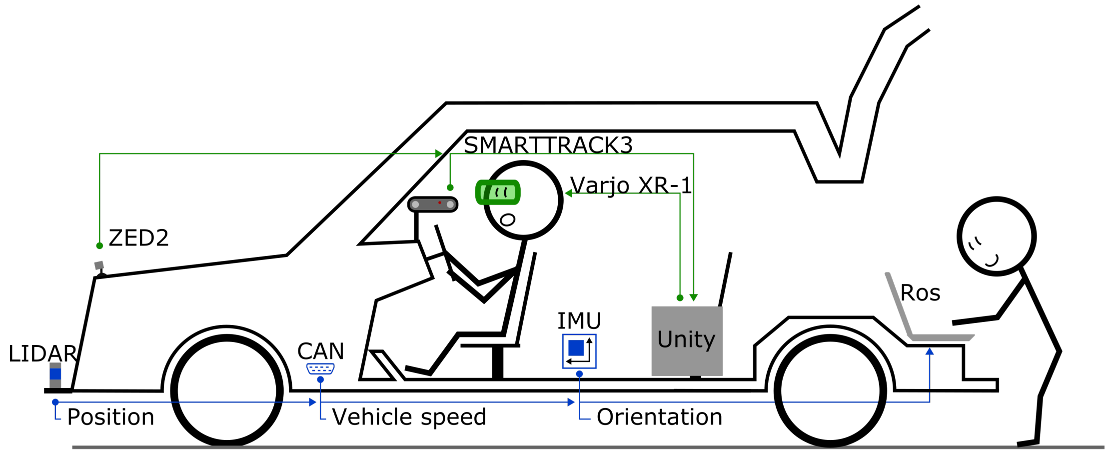


Figure 3.2: Layout of the *XR-OOM* system components and validation equipment in research vehicle, as well as the components used to validate the *XR-OOM* system for this paper.

pant's head in the car, the location of the vehicle in the space, and the location of the virtual objects in the area.

Research Instrumentation

For experimental research purposes, the driving simulation system needs to be able to capture the environment, the participant's perspective, the behavioral response of the participant, the vehicular response from the car, any aspects of the controlled and uncontrolled environment which might influence the study outcomes. We use the eye tracker of the XR-1, a 360° activity camera mounted under the rear-view mirror, and an external activity camera stationed outside the vehicle to capture the surroundings.

This kind of rich scenario and behavior capture must involve quantitative reconstruction of the interaction. To this end, the initial version of the *XR-OOM* simulator generated a significant amount of data; it produced approximately 133MB/s; a 10-minute experiment generates a file size in the 10s of gigabytes in its raw state. In future work, data, such as the raw point-cloud stream, could be simplified by using real-time slam algorithms.

Besides recording the sensors in the vehicle, the attached cameras focus on capturing the participant directly. The current continuous recording limit is about 40 minutes, mainly limited by power supply limits and memory available on the cameras and computers. This could be extended to 2-3 hours with integrated power supply setups in the vehicle and a different storage setup.

Capture Participant Response

There are many ways in which this simulator supports capturing the participant's behavior.

Video By recording the participant's perspective inside the car, the view outside the car, and the surrounding environment, we gained valuable insights into how interactions unfolded and how participants responded to the driving and cockpit tasks. This comprehensive approach enabled us to reconstruct and analyze the complex dynamics of the traffic scenario, providing a deeper understanding of human-automation interactions in the context of autonomous vehicles.

Gaze Tracking Most research-focused VR/XR headsets include an Eye tracker that allows you to record what the participant is looking at. This can help researchers reconstruct what the participant focused on, what they spent time on, and what caught their attention.

Open-ended questionnaires Open-ended questionnaires were employed as a valuable tool in our research to capture the participants' subjective experience. By allowing participants to express their thoughts, feelings, and perceptions freely, we gained insights into their overall experience and well-being during the study. These questionnaires provided a platform for participants to provide detailed and nuanced feedback, helping us understand their perspectives, preferences, and any challenges they encountered. The self-reported data obtained through open-ended questionnaires enriched our understanding of the participant's experience and complemented the objective measures captured through other research methods. We also used the think-aloud method to capture the participant's experience and reasoning directly.

Participant's input Participant behavior was constructed from the video and system recordings. These were carefully monitored and abstracted into measures such as steering wheel movements and mistakes. These measures provided valuable insights into how participants interacted with the augmented environment and the autonomous systems. By analyzing steering wheel inputs, we could assess participants' control actions and asses their certainty in executing the task by counting steering wheel reversals. Similarly, recording participant mistakes provided information on performance errors, allowing us to identify areas of improvement and potential challenges when using the *XR-OOM* system.

3.3.3 Position and Trajectory Measurement

To measure the different patterns between the conditions, we need a system that calculates the trajectory of the vehicle. We used Google's Cartographer [92]to extend a separate set of tracking sensors with a localization system. The trajectory from the vehicle can be extracted to generate performance measures like, e.g., center line deviation. An example of this output can be seen in Figure 3.3. For more details, please see Section 3.4.2.

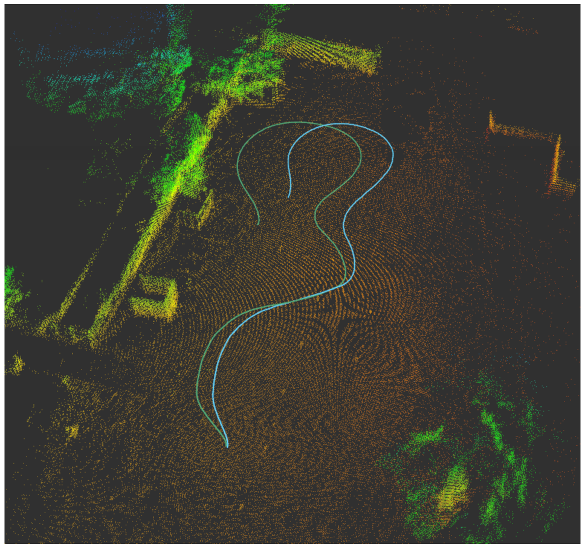


Figure 3.3: Depiction of the path as calculated by the ZED 2i (blue) and the SLAM algorithm (red) overlaid on the depth map from the Ouster LIDAR.

3.4 System Evaluation and Study

We perform a system validation test to evaluate our *XR-OOM* system to benchmark the external validity of simulated activity; this validation measures how similarly participants respond to real-world vs. virtual objects while using *XR-OOM*.

The following system validation focused on operational issues in running mixed reality in-vehicle driving simulations. In addition, we evaluate the safety and comfort when using the system, its effects on driving performance, and its usefulness in understanding and recording driver interactions.

Overall, this validation aims to confirm that the *XR-OOM* platform can be used to run user studies with external validity safely. Initial test drives with the headset worn by a researcher in the passenger seat were used to develop in which a passenger on the vehicle was wearing an XR headset to evaluate the performance and comfort levels associated with wearing the headset.

3.4.1 Pilot Test

In pre-validation pilot testing, we demonstrated that the system's graphic rendering works on an experiential level. The registration between virtual objects and the physically projected world was convincing, such that it matched the physical sensations.

The pilot test involved four different runs with increasing technical complexity. The first pilot had the participant in the passenger seat and reported their experience. The following three pilots then implemented more and more of the final study design, gradually increasing complexity—the pilot participants were members of the same University lab that was not actively working on this project.

The main takeaways from the pilot tests were that **1)** it was possible to acquire situational awareness while using the XR-headset, **2)** it is possible to operate the vehicle in a low-speed environment safely while using the XR-headset, **3)** backward driving required a more complex implementation then anticipated. Navigating backward without properly working mirrors that include virtual objects was not representative of regular driving. We also learned more minor details around the study execution and how to instruct participants best using the setup.

Unlike this pilot test, the subsequent validation test aimed to ensure that the system works for a broader audience of participants and verify that the approach is practical for use in driving simulation studies.

3.4.2 Sensor and Localization Validation

To validate the *XR-OOM* system's ability to track and localize the research vehicle as it moves through the real-world study environment, we set up a secondary trajectory tracking system to benchmark *XR-OOM*'s tracking and localization capabilities. This system also enabled us to track vehicle motion through all study conditions, including the control condition where the *XR-OOM* system was not used.

The trajectory was computed using the Simultaneous Localization and Mapping (SLAM) solution from Koide et al. [118] running in Robot Operating System (ROS) [170]. Below is a list of sensors that were used. For the moment, the LIDAR was used to generate the trajectory and the map while speed and IMU messages were recorded.

LIDAR A OS1-64 2nd Generation LIDAR sensor from Ouster was mounted at the front of the vehicle, running at 1024 beams with 20 HZ. This sensor was used to build a depth map of the environment around the research vehicle.

IMU An Xsense MTI-300 IMU was mounted in the center of the vehicle and connected over USB to the ROS core computer; this enabled tracking of the vehicle's orientation in space.

Odometry Odometry data from our research vehicle, a 2015 Toyota PriusV, was obtained from the vehicle's CAN Bus using a Korlan USB2CAN module. This was primarily used to track the forward velocity of the vehicle; data was sent over USB to the ROS core computer.

Networking The validation stack used the networking infrastructure of the *XR-OOM* system to connect the LIDAR and the SmartTrack3 to the ROS core Computer.

Power We estimate the power consumed by the validation stack to be 250 W, supplied by the same power infrastructure that powered the main *XR-OOM* system.

3.4.3 Validating *XR-OOM* as a Testing Platform

While driving simulators have been widely used for assessing fitness to drive, there is no systematic way to compare their validity and fidelity [23]. Since XR simulators using video pass-through are relatively new concepts, no clear and objective evaluation criteria exist for the impact on driving performance.

To ensure the system's capability to capture driving performance and participant response and to understand the impact of the system on both these factors, we conducted a validation test to ensure that the system meets the criteria for being used to conduct user studies typical of those that human-machine interaction designers and researchers run. The criteria that we investigate are the following:

1. Does using the video pass-through headset allow drivers to conduct cockpit tasks?
2. Does the video pass-through headset allow participants to maintain sufficient driving performance and experience?
3. Are virtual objects simulated believably enough to provide valid predictions of how drivers will respond to physical objects in driving?
4. Does the system support data recording to the extent required to run interaction user studies?

Moreover, all of these functionality checks needed to occur while the vehicle was being safely operated, so this was also a validation of the system's safety for this use case.

For this validation, we had participants perform cockpit tasks and driving tasks with and without the XR headset. Our test version of the *XR-OOM* platform was implemented in a Toyota Prius V, used for the validation tests in a restricted testing area in a parking lot, with permission from the municipal lot owners.

3.4.4 Tasks and Course Design

We developed several basic study designs that an in-vehicle mixed-reality driving simulator could test. The participant's performance of these tasks helped us understand the system's performance envelope, allowing for studies.

3.5 Study

In the following section, WeI will provide an overview of the study, including participant recruitment, methodology, and experimental conditions.

3.5.1 Participants

As our test was focused on system validation rather than experimental inquiry, the validation sample of participants was drawn from our university through convenience sampling. One of the participants' recordings was incomplete and hence dropped from the analysis. The recording of one of the participants is incomplete. Therefore we removed the participant from subsequent analysis. The resulting sample of 10 participants covered the following estimated demographic characteristics: age (20-30), gender (5 male, 5 female, 0 non-binary), and nationality (India, United States). Two participants reported only having driven right-hand drive vehicles before this study.

3.5.2 Method

Experimental Conditions

Our experiment had three conditions:

*Condition **A**. No headset* Participants would not wear a headset when conducting tasks and act as a type of control against which the other conditions can be compared against.

*Condition **B**. Headset with video pass-through only* Participants would wear a headset that would show a pass-through video of the environment but which included no virtual objects. This condition isolates the impact of wearing and using the XR-1.

*Condition **C**. Headset with video pass-through and virtual objects* Participants would wear a headset showing the same view as condition **B** but with virtual objects overlaid. This condition evaluates the participant's ability to see and react to virtual objects.

This was a within-subject study design; all the participants experienced all conditions. We used a Latin Square ordering method to counterbalance the order of the headset conditions for each participant.

Cockpit Tasks

Part of the experiment was to perform several non-driving tasks in the cockpit. These cockpit tasks were designed to test the participant's ability to operate the cockpit interfaces of the car. The cockpit tasks were done two times, once without and once with the headset. They were added to the beginning of the no-headset condition **A** and the first condition with a headset (either condition B or C).

We focus our cockpit task validation on stationary tasks that fit into what Bubb et al. call "secondary tasks," [21] which involve communicating with other traffic participants or reacting to the environment changes. Based on the literature [67, 112, 143, 199], we developed the following list of stationary validation tasks: **Turn on vehicle, Adjust seat, Fasten seat belt, Adjust mirrors, Use turn signals, Control headlights, Control wipers, Control hazard lights, Using the parking brake, Verbally explain which dashboard lights are visible.** The version presented to the participant can be found in Appendix A.2.4.

This is an important validation step for a simulator as there are many cases in which the driver needs to read and react to information from the dashboard or road. A simulator that severely limits the participant's ability to resolve the world around them would be significantly limited.

In this study, the cockpit task was also a safety check before the participant would drive the vehicle. We could externally verify that the participant could operate the vehicle. Both in their ability to operate the vehicle and their ability to see the world through the XR headset.

Driving Tasks

Following the stationary cockpit tasks, we evaluate the mixed-reality in-vehicle driving tasks around the category of Bubb et al.'s "primary tasks" [21] are to keep the vehicle in a planned path and include navigation, handling the road situation in front, and low-level lateral and longitudinal control.

Using a video pass-through headset can introduce driving impairments similar to those of people of advanced age who experience degradation of driving abilities [9]. Their driving impairments could affect the participant in a multitude of ways. These could be cognitive impairments impacting, e.g., their visuospatial ability, speed, and reaction time; vision impairments such as visual acuity, visual

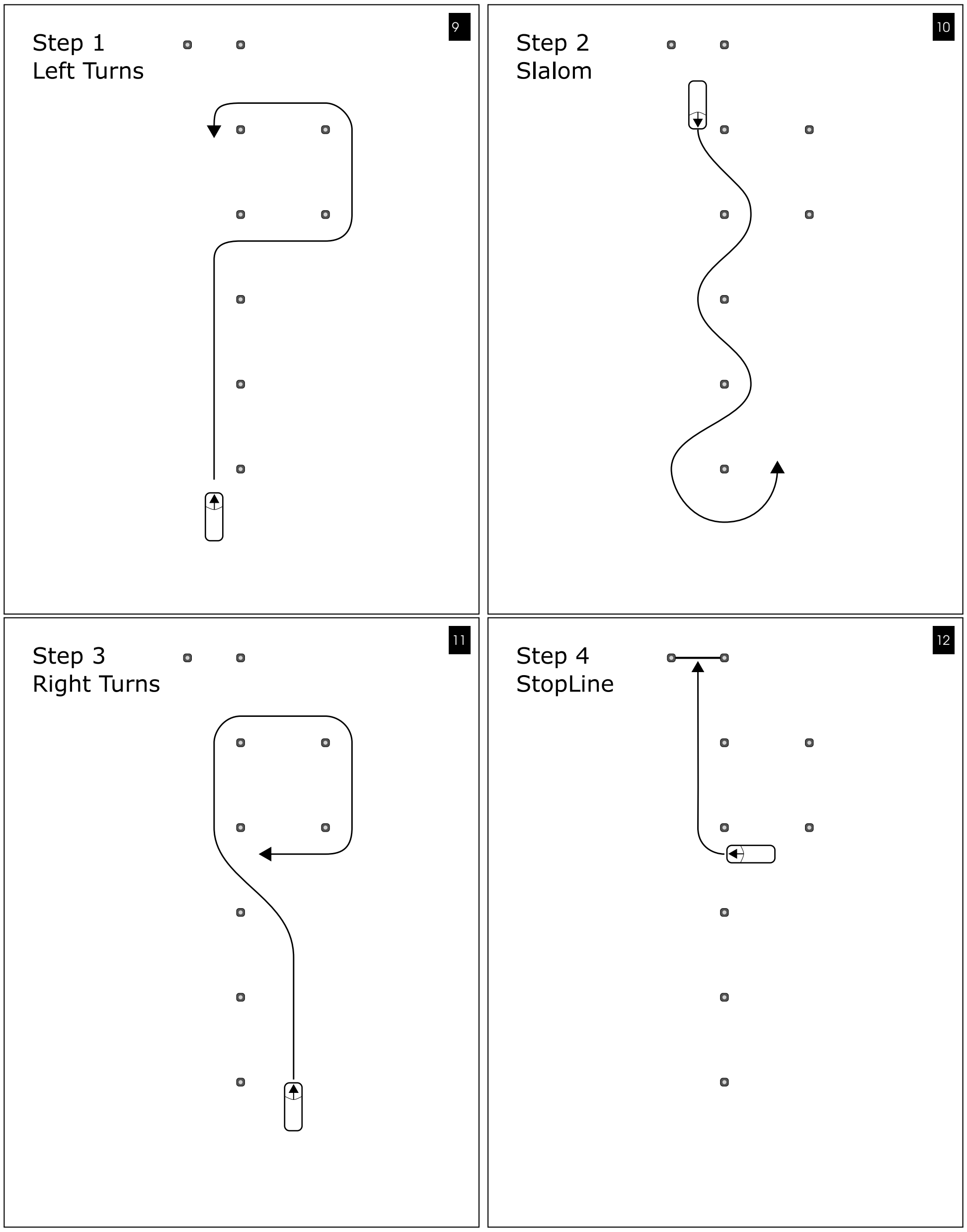


Figure 3.4: The cockpit tasks the participants drove. Each step was printed and sealed on a letter-sized sheet of paper. The dots on the page represent the cones placed in the parking lot.

fields, color vision, depth perception, contrast sensitivity, and glare; and physical impairments such as limited neck rotation and trunk rotation. Our validation activities aim to ensure that the impairment due to the video headset does not unduly affect primary driving capability.

Drawing upon driving task literature [113, 146, 171], as well as resources for people learning to drive in the US [31], we defined four driving tasks: **Left Turns, Slalom, Right Turns, Stop Line** (see Figure 3.4). The tasks took place in a parking lot marked using traffic cones. In the no-headset **(A)** and a headset with video pass-through condition **(B)**, the cones were physical cones placed on the road at 8-meter intervals according to a pre-defined pattern. In the headset with virtual objects condition **(C)**, the physical cones were removed from the lot and visually replaced with virtual cones displayed within the XR headset, set out in the same pattern as the physical cones.

These first evaluation steps evaluate the participant's capability to safely operate the vehicle and perform basic driving and cockpit tasks, forming a driving simulator's baseline requirements.

LT: Left turns
S : Slalom
RT: Right turns
SL: Stop line

Procedure

Each participant's engagement took about an hour in total. The participant was first asked to give informed consent for the study and then was transported from campus to the testing site. The participant was told they could stop the experiment anytime for any reason. They were also informed that they should let us know if they feel motion sick or unwell in any other way. We asked them to drive slowly and that, in the unlikely event of a failure of the mixed reality system, they should stop the vehicle by stepping on the brake. The participants performed all the study tasks as ordered by the study's Latin Square design. After the driving tasks, the participants answered simulator sickness and subjective experience questions. After the experiment, the experimenters would transport the participant back to campus.

Measures

We used a variety of observed, objective, and self-assessment measures to understand the *XR-OOM* system's impact on people's driving.

Observed measures. Video of participant performance in the cockpit and driving tasks were observed and rated by researchers post-facto.

Objective measures. The driving performance was measured using post-facto analysis by computing the Frećhet distance between a normalized driven path and a normalized "ideal" synthetic ground-truth path. This comparison of the driving trajectories, detailed in Appendix A.2.5, was visually confirmed against the 360°-video recordings of participants, particularly in cases where the participants were clearly far off the ideal path.

Self-reported Assessment We performed semi-structured interviews with participants to record their own impression of their ability to drive with the *XR-OOM* system. We asked people to describe their experience driving with the XR system and note differences between with and without it.

3.5.3 Risk Assessment & Mitigation

Real-world experiments contain the risk of physical injury to the participant and other people and the risk of physical damage to the vehicle and the physical world. Our broader goal with the *XR-OOM* system is to run on real roads with regular traffic. Still, for this validation study, we exercised an abundance of caution to minimize possible sources of risk. A list of risk factors and our mitigation methods is given below:

Driver Variability

We recruited drivers with driver's licenses and were legally allowed to operate vehicles in the United States. The pre-study protocol involves simple instructions for gaining consent and setting up driving activity. Participants who had difficulties with these steps would not have been asked to continue with the study. We could have further restricted participation to people with local driving experience and experience with automatic transmission vehicles; these restrictions trade-off against understanding issues that a broader diversity of participants would reveal, so we did not apply them to this study.

Unintended Acceleration/Collision with Obstacles

We designed the activities to occur at low speeds and over short distances. We located the tasks so that strong braking or accidental acceleration would unlikely result in a collision with any surrounding structures. The participants were asked to put on seatbelts early in the study, and those instructions were repeated at the start of every condition. We tested both physical and CAN-bus-based systems to limit the vehicle's speed. Subsequently, We found they were unnecessary and could introduce other safety issues because they could cause the car to behave in ways that participants would not expect.

Collision/Interference with Other Vehicles and Pedestrians

We set our study in a restricted lot where we would not encounter wayward traffic. We had a permit for the exclusive use of the lot during our studies. The on-site researchers also directed people and vehicles driving through the site so they did not come close to our test area. In advance, drivers were told that people drive through the site to just brake and wait until the other vehicles had passed through.

Researcher Safety

The study was initially designed to occur with a researcher in the backseat of the vehicle the participant was driving. To reduce the risk of viral transmission, we subsequently located the researcher outside the vehicle and provided participants with instructions and guidance over a mobile phone. Researchers ensured participants were entirely stopped and asked them to put the car in park before approaching the vehicle and opening any doors. Researchers wore reflective traffic safety vests to increase their visibility and stood at the same location throughout the study.

Loss of Vision

Because the system uses a video pass-through, it is possible for hard failures in the XR system to cut driver visibility completely. In any on-road experiment, it is also possible that external obstacles and events can cause loss of vision. We told participants that in the unlikely case of complete loss of the ability to see the vehicle and driving site, they should immediately step on the brakes for a hard stop and await additional instructions. The loss of vision could come from **Hardware Failure** like a loss of power or a cable disconnect. We ensure the system's safety through repeated testing and piloting, as well as using warning systems for the battery power level. **Software Failure** was another possibility that could lead to a loss of vision. This could be triggered by Unity failing, the graphic card drivers, or even the entire system. This was mainly mitigated by ensuring we frequently restarted the system and only ran the required software.

Loss of Tracking

Each of the two tracking systems **SMARTTRACK3** and **ZED 2i** could fail to break or interrupt the continuous tracking of the motion of the headset through the world. The main problem was dealing with differing lighting conditions. Piloting the system in different/acceptable weather conditions helped us ensure that loss of tracking was not a frequent occurrence and that the system could automatically recover from it.

Driver Discomfort

Discomfort is a factor in many driving simulation studies, so much so that there are standardized questionnaires to assess the degree of "simulator sickness" participants experience [111]. Counter-intuitively, both Paredes et al. and Goedicke et al. found that, with a high-resolution virtual reality headset in a vehicle with well-correlated physical and virtual motion, the combination of the virtual reality environment and the physical environment causes *less* motion sickness than virtual reality environments or in-vehicle autonomous driving experiments themselves [74, 167]. This is probably because the correlation of virtual and physical motion induces fewer problems than the juxtaposition of moving environments without corresponding visual movement (as commonly occurs in VR environments) or visually moving environments without corresponding physical motion (as occurs in driving simulation) [100].

While we would like not to have any discomfort, we wish to understand whether any experienced discomfort is tolerable or impairing. To minimize the likelihood of nausea and motion sickness, we had the air conditioning turned on in the car with the fan speed set to high. We also let the participants know they could stop the study at any time if they felt uncomfortable. With mixed reality in a moving car, both motion sickness and simulator sickness need to be avoided [187]. Participant discomfort caused by simulation or motion sickness could further be mitigated by good ventilation and temperature control, which is a feature that can be expected in most on-road vehicles [187].

We did not include tasks that would involve reversing; the entire experiment takes place in the forward-driving direction. This decision was based on two considerations: (1) rear-driving tasks are already more challenging in normal driving conditions and thus include a higher risk, and (2) our system scope currently does not support a whole rear driving experience due to virtual objects not appearing in mirrors inside the vehicle.

3.6 Results

3.6.1 Task Measures

In the following sections, we discuss how participants can control the system to validate the *XR-OOM* system's usability as a simulation platform.

Cockpit Tasks

	With Headset		Without Headset	
	Completed	Failed	Completed	Failed
Turn on vehicle*	3	2	6	0
Adjust seat*	5	0	5	0
Fasten seat belt*	3	0	6	0
Adjust mirrors*	9	0	7	0
Use turn signal	10	0	8	1
Control headlights	10	0	8	0
Control wipers	9	1	9	0
Control hazard lights	10	0	9	0
Use the parking brake	7	3	7	2

Table 3.1: This table shows the performance for different cockpit tasks expressed by counts of completed and failed tasks across participants. (* In some cases, a task was skipped because it was already completed, e.g., the seat was already in the correct position. This leads to the total count amounting to less than the number of participants (10).) The *Explain dashboard lights* task results are excluded here because the counts do not indicate when participants erred.

Cockpit task performance was evaluated as either success or failure. The data was extracted from the 360°camera in the same manner as for the driving tasks (the entire table is in the Appendix 3.1. Most participants could successfully conduct the tasks with minor retries in the headset condition. Some participants could not turn on the vehicle (N=2), used the wrong stalk for controlling the wipers (N=1), or looked for a button instead of operating the foot-controlled parking brake (N=3) in the headset condition. The cause of these errors was mainly related to unfamiliarity with the (type of) vehicle or the instructions, such as not knowing to press the brake when starting the vehicle or not knowing that the parking brake was foot-operated. The failed tasks did not rely heavily on visual information. In five tasks, some participants skipped the task due to the task already being completed (e.g., the seatbelt was fastened) - we did not include their counts in the total number of failed or completed tasks.

One notable exception was the last task in which the participant had to explain all the icons they could see on the dashboard. The vehicle speed (the largest number) was quickly visible; however, smaller icons (e.g., tire pressure warning) could not be resolved. While some participants could see more than others, they could only see and report some icons. However, most participants were able to read the speed indicator, which is the largest number on the screen.

Many participants also completed the parking brake task wrong and pressed the [P] button instead of the foot operating the parking brake, which was the intent for this task. Arguably, however, this was due to the ambiguity in the instructions and had nothing to do with the headset affecting the participant to do the task. This is based on the fact that failure to complete this task was similar across all conditions.

One notable difference during the study was that most participants needed to

	A				B				C			
	LT	S	RT	SL	LT	S	RT	SL	LT	S	RT	SL
Correct path*	8/8	6/8	9/9	9/9	7/8	5/9	8/9	9/9	2/9	3/7	4/9	5/9
Halted	1/9	5/9	3/9	1/9	1/9	6/9	2/9	2/9	6/9	4/9	3/9	1/9
Task time M (s)	41.9	70.2	42.2	17.3	43.7	57	46.4	15	50.9	49.4	54.7	26.7
Task time SD (s)	17.7	41.6	16	6.8	27.3	33.5	12.9	6.3	14.8	21.7	11.9	24.2

Table 3.2: Summary of participants' behaviors. For each task and condition, we report the total number who followed the correct path, halted, and mean task time and its standard deviation for task and condition. The abbreviation **LT S RT SL** refer to the four different driving tasks. See Figure 3.4 for more details.
*In some cases, participant results were excluded as they were caused by a system failure (e.g., tracking error).

raise and hold the instruction papers in front of the headset, contrary to reading directly from their lap.

Driving Tasks

All participants could read the instruction diagrams and follow the routes. For the remaining operational driving tasks, their performance was evaluated using the metrics shown in Table 3.2. A researcher annotated the video recordings by observing videos of the participant-facing camera, an external camera, and the headset camera. The annotation criteria of each metric were as described below. A human annotator extracted performance data about the participants driving by observing videos of the participant-facing camera, an external camera, and the headset camera. From these annotations, a set of tasks appeared to be particularly difficult. These are described below. The annotation criteria of each metric were also described below.

The slalom task was the most difficult for participants to accomplish in the driving tasks. Most participants needed to back up to do it correctly because the first turn into the slalom was very tight. This task was unforgiving; if participants missed the first turn, they would have to retrace their position to be correctly positioned for the subsequent task. The short height of the traffic cones made it challenging to see the cones when drivers got closer to them; participants had to remember the locations of the cones and guess where the vehicle's bumpers were in space to avoid hitting the cones. This task was particularly challenging in condition **C** because the lack of ability to see the virtual cones in the rear-view mirror made it challenging to locate the relative location of those cones to the car. Furthermore, this task was challenging because the cones very quickly vanish below the vehicle, and you need to have a good guess of where the vehicle starts and ends to avoid hitting the cones.

For the *Left Turns* task, finding the correct point to turn right was also hard for some participants. Several participants turned too early and then had a very long left turn.

Participants found the Stop Line task to be the easiest. However, the participant's interpretation of where to stop for a stop sign line varied significantly.

3.6.2 Driving Trajectory

Each participant's drive was recorded with ROS. However, two trajectory recordings were incomplete or corrupted after the experiment, so the data for 8 participants could be compared. From 8 recorded trajectories, we compute the Fréchet distances as a proxy for driving performance to compare the conditions against an "optimal" path. Please see Appendix A.2.5 how these values were computed. The graph of the performance differences from these analyses (see Figure 3.5) The average Fréchet distance increase from condition **A** to **C**, though not

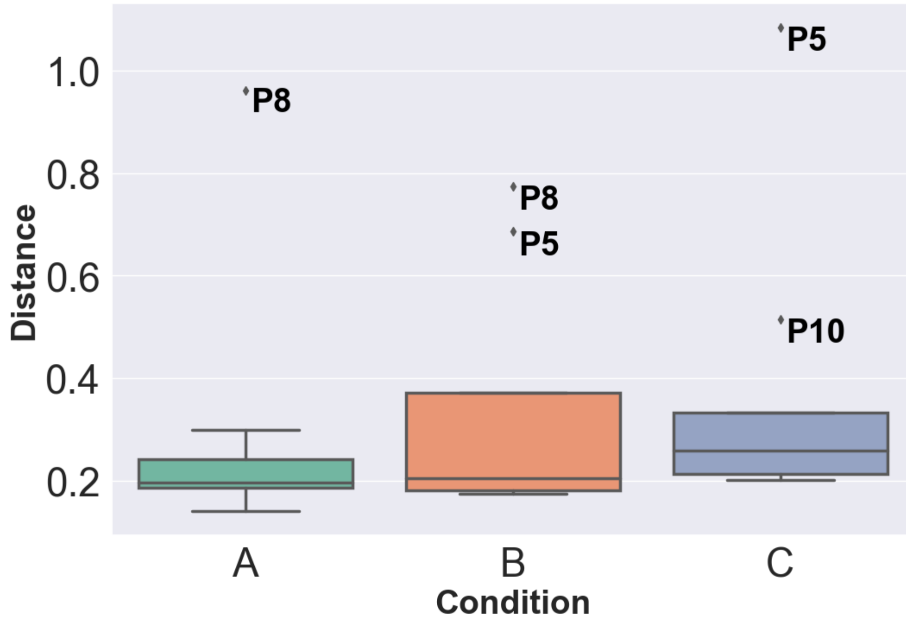


Figure 3.5: This box plot shows for each condition the Fréchet distance with the outliers showing the participants and conditions where the participants miss judged their path and followed the wrong route.

significantly different with 1 to 2 outliers for each condition.

3.6.3 Self-reported Measures

We also recorded each participant's post-experiment interview results, including a self-assessment rating for each condition.

At the end of the experiment, participants were asked eight questions about their general experience and well-being. Three out of 10 participants mentioned they drove better *with* the headset. Three participants reported mild motion sickness, one of which was more severe. Three people reported a light headache, and 4 participants reported some eye strain. Although participants were asked to contact us if these symptoms failed to subside or if they got worse later in the day, no participant contacted us about these issues afterward.

3.7 Discussion

As a driving simulation platform, *XR-OOM* must support a wide range of experimental studies. With the results from this implementation, we can now validate the system and contextualize it with other types of simulator studies and possible scenarios this kind of system should ideally support.

3.7.1 System Validation

We return to the questions first raised in Section 3.4.3 for system validation. We should point out at the outset of this discussion that since we are trying to validate the XR HMD setup for driving studies, the desired result is for there to be as little difference as possible between the performance between conditions. If there is a difference, we would prefer that difference to be small and the performance to be best in the no-headset condition. This would suggest that tasks people can perform well enough in the XR setting would predict even better performance from the simulation. (This is different from what would be desirable in an experimental study, where low or no difference between conditions would indicate a null result.)

Does using the video pass-through headset impact the participants' ability and experience to operate cockpit tasks?

Besides the cockpit task, we intentionally chose to have the instructions printed on paper for the participant to read to ensure a correctly working HMD and evaluate the participant's ability to read and follow the instructions. All participants were capable of reading the cockpit tasks from the instruction sheet. So most of the cockpit tasks were also unaffected by the use of a headset (see Figure 3.4). These results point to the system's exciting capability that allows for testing of brought-in devices like phones or tablets during the driving experience. These devices might be the participants' real devices or virtual applications that do not yet exist.

The most problematic task was distinguishing icons on the dashboard. Cameras resolution, dynamic range, and refresh rate aliasing made this task particularly difficult. While all participants could see some information, most only reported the vehicle speed, the largest number in the screen center. Additionally, it is essential to point out that seeing the media center and AC controls caused fewer difficulties; however, these were not explicitly part of our study and need further assessment in future research.

Certain display/camera interference notwithstanding, this method allows for testing new displays and controls. With the present technology, it is possible to change any display surface. These should be done in a way that has face validity (simulated interfaces should strongly resemble actual interfaces) and elicit relatively realistic behavioral responses. This also includes heads-up displays and other new display technologies that go beyond current instrument clusters or center console media players.

Additional complications occurred with the tasks that required body movement in the vehicle, like setting up a chair and putting on the seat belt. Wearing a headset gave less space for the participant to move around, and the limited field of view from the headset meant that many steps in these tasks needed to be done via touch (e.g., feeling where the seat belt is).

The cockpit task did not pose any significant hurdle to running the study. For such tasks, further optimizations like directly controlling the XR camera's exposure could help make display items more visible. To address this issue, future work could look at how a subroutine could adjust the headset's exposure based on its pitch angle.

Does using the video pass-through headset impact the participants' driving performance and experience?

Based on the experiments, the difference between driving with a XR headset vs. without one is minor. Variables like average speed and task time are similar, and the vehicle paths appear similar when visually compared.

The main limitation identified during these trial runs was that the headset's weight and bulkiness caused participant discomfort. This, in combination with the limited space of the vehicle cabin, detracted from the participant experience.

These findings pertain specifically to low-speed driving tasks; we expect that driver performance and the driver's experience would degrade in higher-speed conditions. However, based on the experiences we have had to date, it should be possible to safely perform driving tasks up to 30 mph if the vehicle is in an area without other cars or pedestrians.

Even with the mentioned limitations, these findings enable this approach to test drivers' responses to different visual impairments to the driving environment (e.g., cataracts, heavy snow, or rain) and to design compensatory measures for such drivers. By altering the visual feed to the XR headset, we can simulate the impact of visual impairments as limited field-of-view have on driving.

Are virtual objects simulated believably enough to provide valid predictions of how drivers will respond to physical objects in driving?

One key issue for this validity study is whether participants treat the virtual objects as real objects enough so that their interactions with the virtual object indicate how they would interact with future physical objects. The realism of the simulation could be related to aspects such as depth perception in virtual displays, which has been studied in various automotive AR applications such as [20].

From the study footage, we could see that participants drove more cautiously through the tasks with the virtual cones, taking a little more time and making a few more mistakes. In general, completing the tasks with condition C was more difficult than any of the other conditions. This is also reflected in the participants' self-reported experiences.

Analysis of the three conditions' driving paths (Figure 3.5) shows that people tended to drive similar paths in all three conditions. This also indicates that the participants did not ignore or in any other way treat the virtual cones any differently than the physical cones. After watching the headset video recordings and examining the generated path, the participants seem to take wider turns around the virtual cones, comparatively showing more cautiousness when interacting with virtual objects.

This suggests that participants took the virtual objects seriously and tried not to hit the objects even though they were not physically present.

For future studies, the traffic cones could be replaced by pedestrians, bicyclists, road incursions, road signs to be followed, etc. Additional XR elements could then be introduced into the XR environment to accentuate the external phenomena (e.g., highlighting pedestrians it senses and recognizes) or feature simulated objects or actors to assess the participant's response to the highlights or simulated object.

The capability to meaningfully place and track objects in the environment also extends the capability of the simulator towards testing extraordinary traffic conditions like, e.g., an icy patch of road that affects user experience and vehicle response. While the current system cannot affect the control behavior of the car, it can visually augment the participant view to create a more holistic experience in ways that a controlled laboratory or pure VR studies would not be able to.

Does the system capture information about participant behavior well enough for analysis in user studies?

In the analysis above, we highlighted some of the available data streams we captured and demonstrated how they could be used to gain insight into a participant's behavior and performance.

The *XR-OOM* system offers multiple perspectives for qualitative coding that researchers can use to observe a participant post-facto. Video streams provide a clear view of the vehicle, the participant, and the road in front of the vehicle and directly show the participant's perspective and track their gaze. These streams can also be automatically analyzed, as shown in Figure 6.2. These different perspectives allow the researcher to reconstruct the complete picture of what the participant saw and how they reacted in a given situation.

Capturing detailed participants' responses is key when trying to understand their reaction to a given social environment; these have been shown to have an enormous impact on both good and bad drivers. The combination of *XR-OOM* capabilities to collect detailed data and render virtual objects allows for simulated social actors in the car. By adding sound sources and other virtual characters into the vehicle, we could, e.g., evaluate how social distractions affect a driver and how technology could help mitigate such effects. Conversely, social agents could prevent drivers from becoming bored or angry in the vehicle or help point out important external cues to improve situation awareness.

Further, qualitative data can be gathered through self-reported questionnaires and interviews. This study used a regular phone connection to allow participants to think aloud as they worked through the different tasks and answered questionnaires at the end of each condition. This is made possible by the headset's high resolution as it resolves the phone screen enough that the participant can use their phone.

The quantitative data allows a more formative analysis of the participants' driving. Sensors like the used LIDAR or IMU will enable us to capture the detailed movements of the research vehicle and compute an accurate path that can highlight the driving behavior.

The combination of participant behavior and vehicle motion are vital features to test advances in vehicle automation. *XR-OOM* could support the study and development of different paths to driving automation, such as teleoperation [159] or supervisory control [98]. For such studies, relative behavioral validity in how people notice and respond to control transition events is desirable [94, 158].

Other Discovered Issues

Beyond answering the questions we posed about the viability of *XR-OOM* as a driving simulation environment, our research also helped us discover technical issues that should be addressed, particularly regarding the cameras used on the XR headset.

Because XR equipment is often designed to operate indoors and under controlled lighting conditions, the cameras on these systems are not well-optimized to handle the high-dynamic range of light intensity in an outdoor setting. This can especially be problematic inside the research vehicle, as participants might switch between looking at the outside world and a shaded dashboard. This use also extends to visual tracking systems, as some depend on their own modulated IR light sources. The sun can quickly “down out” these sources making it impossible for the camera to pick up the modulation and tracking markers.

Another issue we encountered was the interaction between the refresh rates of the camera and the instrument panels the participant was trying to read from. If these have similar refresh rates, aliasing can occur; this makes the panel displays appear to fade in and out and thus become harder to read.

3.7.2 Safety Assessment and Risk Mitigation

For safety assessment and risk mitigation, we return to the previously identified risks in Section 3.5.3. We also discuss the necessity and efficacy of the measures we took for risk reduction in this validation work.

Driver Variability

The participants had different comfort levels and capabilities in accomplishing the cockpit and driving tasks, even for the control condition tasks. Some participants were flustered and inadvertently skipped tasks. Another source of variability came from the participants’ experience driving a vehicle like the research vehicle, e.g., some participants had only driven manual transmission vehicles and therefore took longer to find and use the parking brake. The participants also had differing responses to the XR headset: their differences in height and neck strength affected their response to the system. While we were prepared to reject participants who seemed like they might not perform the study tasks safely, we did not find that we needed to stop the study early for any of the participants in this study.

Unintended Acceleration/Collision with Obstacles

The task design successfully limited the participant's speed as the recorded maximum was only 4.3 m/s. At no point did any safety systems like speed limiters, verbal collision warnings, or safety belts engage.

Collision/interference with Other Vehicles and Pedestrians

In some cases, other vehicles or pedestrians drove through the lot. At no point did it seem like the participants were likely to collide with any other vehicles or pedestrians; additionally, the participant either noticed the other vehicles or passers-by themselves or was notified by the researcher over the audio channel to pause study driving activity. For the duration of anyone passing, the participant halted the task.

Researcher Safety

The researchers had a clear spot to the side of the experiment area. They used clear communication to confirm parking with the participants when they approached the vehicle between conditions.

Loss of Vision

Because the system uses a video pass-through, it is possible for hard failures in the XR system to cut driver visibility completely. These risk factors are discussed in Section 3.5.3. While we did not have any complete HMD blackouts, we did have an instance where the HMD popped dialog boxes on the display, which blocked the areas of the participant's visual field. In that instance, the driver could stop and follow audio instructions to resolve the issue. In the future, finding ways to turn off such messages is important unless the vehicle is completely stopped. However, the possibility of this type of failure leads us to recommend that driving with HMD should be restricted to low-speed ($\leq 15mph$) driving scenarios in real-world environments and mid-range speeds ($\leq 30mph$) in closed track or black-lake environments. Higher speed driving should employ XR systems with optical pass-through, such as the Microsoft Hololens [148].

Loss of Tracking

Situations where the headset's location within the car was not correctly mapped, occurred several times during the study. We noticed that when some participants turned their heads to the left, towards the driver's seat window, the in-car localization lost track of the headset. This was not a significant problem as tracking was resumed as soon as the participant turned their head back.

The in-world tracking was also unstable for some participants; in particular, variability in external lighting caused by cloud cover caused the tracking to be less reliable. The participants were always able to continue the experiment. The tracking loss resulted in small jumps (a few centimeters and degrees) in the visual field; no participants that experienced this tracking jitter had difficulty adjusting their trajectory.

On the whole, and comparing the technical issues using this simulator to previous in-lab and in-vehicle driving simulation environments used by the study researchers, we felt this system was robust enough to perform driving experiments with.

Driver Discomfort

While we anticipated and hoped to mitigate discomfort stemming from the simulator or motion sickness, we found that the major source of driver discomfort in the study came from the headset, which many participants found to be heavy.

Prior research has indicated that heavy or unbalanced headsets can create discomfort in participants over time [114, 213]. We hope future headsets are lighter and less likely to cause this issue.

Across the study conditions, condition C was rated the least comfortable but was still above the mid-point for the Likert (2.6 compared to 2.5 out of 5). We also asked participants to email us if they felt nausea later in the day following the study, and none of the participants sent an email indicating post-experiment discomfort.

3.7.3 Limitations

COVID-19 Protocol Measures

While this system was developed before the COVID-19 pandemic, it was evaluated during the period following strict social distancing measures. Our study protocol included specific transmission prevention measures not detailed above: Participants and researchers must wear face masks throughout the study. The headset and research vehicle were extensively sanitized before and after each participant run. The research vehicle window was left open to enable extra ventilation. Participant recruitment was limited to the on-campus population who were part of an extensive university testing program.

We do not indicate that these preventative measures impacted the study results, and all of these measures should have affected all of our study conditions equally.

Convenience Sampling

Due to the study above restrictions, our study participant pool was limited to part of an on-campus population. Ideally, this work would be replicated with a broader demographic pool to see if age or other factors affect the system function.

Reverse Driving Tasks

Participants in our study could conduct short stints of backing up the vehicle but, on the whole, were not asked to perform reverse driving tasks. Because driving in reverse would require a model of the world behind the vehicle, instrumentation of the rear-view mirror, and complex graphical modeling, this system was not validated for tasks that would require extensive backward driving, such as parallel parking maneuvers. This is a limitation common to many driving simulators. The scope of the current *XR-OOM* system is focused on forward-driving tasks.

Other Road Users

This current paper's contribution focuses on the system design and validation of the safety and usability of the system for driving simulation experiments. This was critical to establish before using *XR-OOM* in environments with other drivers, bicyclists, and pedestrians. Additional safety and validation research should be done before using this system for driving simulation in an environment with other road users.

Order and Carryover Effects

Our experiments discovered carry-over effects in the driving conditions (A without headset, B with headset with pass-through video, and C with virtual headset objects).

Specifically, we saw that tasks A and C are ranked lower across all metrics when task A is run first. When task A is run after task B or C, it is rated higher across all metrics. Task B is rated higher when done after task C, and task C is ranked higher in control when done first. Similar effects can be seen when we analyze the impact of the second task run. In the last task analysis, all tasks are rated lower across metrics when task C is the last task performed. This implies that other tasks are perceived as more accessible after completing task C, and task C is perceived as more complex when run after the other tasks.

This is not a concern for our system validation test or for experiments that take place only in a headset condition but should be kept in mind when the *XR-OOM* system is used for comparing different headset conditions.

Example deployments

We recorded video prototypes showcasing different ideas to highlight what kind of studies *XR-OOM* can support. Here we will show screenshots from those prototypes and explain their background.

Figure 3.6: The four images show different applications in which *XR-OOM* could be used.
The *Top-left* Wizarding autonomy with a virtual driver. *Top-right* simulating on-road events. *Bottom-left* Virtual in-vehicle displays and interfaces.*Bottom-right* Example of situational awareness test .

The *XR-OOM* system offers various opportunities for extension and enhance-

Figure 3.7: Work by Bu et al. [53] demonstrating their use of *XR-OOM* in tandem with the stationary simulator of the original study.

ment. These include incorporating advanced rendering techniques, implementing realistic reflections, automatically adjusting the position of the virtual sun based on weather conditions, and enabling geo-location features. An important step in further developing *XR-OOM* is integrating globalized or pre-mapped environments into the XR software tracking algorithm. This integration, as demonstrated by Bu et al. [53], provides a more controlled and precise environment for conducting studies. Additionally, a comparative study building on prior research showcases the usability and effectiveness of the *XR-OOM* system. The findings of this study contribute to the growing *XR-OOM* repository and highlight its potential as a valuable research tool. (See Figure 3.7)

3.7.4 Follow Up Work

Moving Beyond Cars

In summary, the *XR-OOM* system has made significant contributions to human-automation interaction and autonomous systems. *XR-OOM* has enabled researchers to study and understand the complexities of AV interactions, trust-building, and collaboration by providing a rich and immersive environment.

The integration of the ROS system and advanced XR rendering capabilities makes *XR-OOM* an exceptionally versatile and compelling test platform, extending its application beyond the automotive and AV domain. This platform holds the potential for evaluating robotic behavior in virtual environments, offering valuable insights to designers. The seamless integration with ROS facilitates rapid iteration and development of designs, making *XR-OOM* and XR methods a useful tool for exploring and experiencing virtual robotics in real-world settings.

3.8 Conclusion

In this chapter, I discussed the work done on the *XR-OOM* driving simulation platform and demonstrated its potential as a valuable research tool for conducting driving studies. The system could support various experimental studies, including testing new displays and controls, simulating visual impairments, and studying driver behavior and performance. The system's capability to meaningfully place and track virtual objects in the environment extends its potential to test extraordinary traffic conditions and assess drivers' responses to different scenarios. The structure of *XR-OOM* allows for flexible addition of these and other scenarios, devices, controls, and simulated AV behaviors.

XR-OOM built on early findings of *VR-OOM* and deployed absolute-referenced sensors systems for the head and vehicle tracking. At the same time, this work highlighted some critical limitations of XR technology; the cameras used by the XR headset need to adjust their exposure quickly in high dynamic range environments and support variable refresh rate capture to avoid refresh-rate aliasing. Making these settings available through software is essential to address a responsive experience. These elements are critical for the safe deployment of *XR-OOM*, especially as faster scenarios are deployed.

Some additional contributions of *XR-OOM* are:

Logistic and SOP Preparing *XR-OOM* for deployment requires significant preparation, as its deployment is in the natural world. This includes getting permission to use parking lots, checking weather forecasts, charging batteries, etc. The goal of the publication of *XR-OOM*'s SOP is to support the replication of this method.

Safe deployment The deployment of especially low-speed tasks can be done safely and directly.

Impact of XR-headsets The experiments showed that while the freedom of motion is somewhat limited by wearing the XR-headset, driving, following traffic rules, and planning trajectories is still possible with contemporary(2022) hardware. This also applies to operating the control surfaces of a vehicle, which participants mostly completed.

Navigating around virtual objects While participants drove more cautiously around the virtual objects, their behavior was still comparable to navigating the physical world obstacles.

Rich streams of data The synchronized data capture setup of *XR-OOM* enables qualitative and quantitative analysis methods. In this study, we coded participants' ability to drive(qualitative) and performed quantitative analysis on their driving path by calculating the Frećhet distance from an ideal track.

Future research could explore integrating advanced rendering techniques, implementing realistic reflections, and incorporating globalized or pre-mapped environments for more controlled and precise tracking within the environment.

CHAPTER 4

INTERACTING WITH OTHERS

Figure 4.1: *StrangeLand* uses a multi-person virtual reality driving simulation environment to help illuminate how drivers interact in different cultures. The participants are wearing VR headsets with leap motion (hand-tracking device) mounted on the front. Their hands are on the steering wheels, and gesturing at each other. Across from them are laptops facing the researchers.

Interaction with other people is highly culturally dependent, part of what defines a culture. Fundamental interactions like greeting one another, signaling to let someone pass, and many other small and large interactions are informed by local culture and norms derived from that. Tourists and their behaviors will often give light to these rules as they unintentionally break them by being unfamiliar with local norms. For example, in NYC, it is customary *NOT* to stand in the middle of the sidewalk, even if it is an excellent spot to take a picture. Breaking this rule makes it easy to detect tourists. Understating these local norms is crucial when introducing autonomous agents into these environments, especially when assessing corner cases.[1]

[1]This chapter reuses material from the original publications [77]. Those works were co-authored with *Carmel Zolkov, Natalie Friedman, Talia Wise, Navit Klein, Avi Parush, and Wendy Ju*, but I was the lead author of those papers and the primary researcher on that body of work.

Multi-user Virtual Reality(VR) simulator could offer an excellent opportunity to uncover some of these natural patterns of interaction. All interaction elements are simulated and can be recorded for later behavioral analysis. Integrating this with simulated agents allows it to test new designs of robots and their behaviors in a simulated multi-user environment to understand how the design changes affect interaction, use of the robot, and ideal uncovers serendipitous interactions with these agents.

This kind of work can also already happen before robots are deployed. A specialized multi-user simulation could be used to look at naturalistic interaction in a given scenario to discover social signals and common interaction patterns that artificial agents or robots might need to adhere to in the future. One such environment is the traffic environment with local and regional differences in driving styles. The understanding will need to develop beyond anecdotes and observation toward a concrete understanding of the parameters, behaviors, and interaction patterns that differentiate driving in one locale from another. Today, autonomous vehicles (AVs) are programmed to drive within the boundaries of the law. Still, they seem to elicit higher-than-normal [180] rates of accidents because they do not conform to local driving norms [190]. Officially, many of these accidents are classified as the fault of the non-AV car [119], but it would be better if the AVs could avoid accidents and faults by adapting. For example, AVs could adapt to how different cultures interpret speed limits, how long they wait or how they slow at a stop sign or before a left turn, what acceptable follow distances are on a highway, and how much room they give a pedestrian.

Vehicle-to-vehicle interaction is a great first scenario for uncovering local interaction norms. While technically, particular interactions are codified in traffic regulations, e.g., the use of indicator lights, we can see differences in how these rules are interpreted and how they play out in the local driving culture and regional context. Hence, this project explores a shared multi-participant VR simulator that can be used to uncover and analyze local behavioral norms, specifically focusing on ambiguous traffic situations that force drivers to interact.

In discussing Uber's drivelers car experiment, Uber's Engineering Director, Raffi Krikorian, stated, "If we can drive in Pittsburgh, we can drive anywhere." [88] This statement was intended to highlight the benefit of testing cars in an environment with poor roads and varied weather. Still, anyone who has driven across borders knows that driving culture varies profoundly from one locale to the next.

Over two decades ago, Oskar Juhlin noted that in designing automated driving, "it is essential to understand how drivers themselves achieve coordination. Computers, running by rules or algorithms, must function together with other road users. They must adapt to them, or the drivers will have to adapt to the new machines. If the artificial drivers are socially incompetent, this could lead to ambiguity and misunderstandings, which seriously strains other road users." [106]

While cross-cultural differences in driving are widely known and accepted, there is limited prior work documenting and detailing these differences, none doing so in quantitative ways that could guide machine recognition or response. We began our design effort by looking at related work that measures driver behavior and captures the interaction between drivers. We explain how these developments informed but also necessitated the StrangeLand system.

This chapter reuses material from the original publications [77].

4.1 *StrangeLand* Overview

For AV to adapt to local norms in human driving, it is critical to profile how human driving differs across geographical locations. While ethnographers have qualitatively described regional differences in driving style, data-driven statistical models help *computer-driven* cars drive like locals and recognize how local drivers are signaling through hand/body movement and motion of their vehicles. To this end, we have created an experimental system (see Figure 4.1) and method to profile driving behavior and interaction using a multi-participant virtual reality (VR) driving simulation environment. The system was designed to be portable and to support cross-cultural experimental deployments. We aim to ensure the system is operational and functional, can model diverse scenarios, generates data fit for analysis, and captures expected behaviors. We describe the system, test scenarios, and findings of the proof-of-concept study conducted in the U.S. and Israel.

4.1.1 Background

Self-report Based Studies

Transportation researchers have used questionnaire- or log-based assessments to profile several characteristic differences between drivers in different regions, often to account for differences in accident rates. Özcan et al. examined differences in driving behavior across six countries–Finland, Great Britain, Greece, Iran, the Netherlands, and Turkey [165]. Using a driver behavior questionnaire [172], The researchers found that self-reported differences in aggressive driving violations, ordinary driving violations, and driving errors corresponded with differences in the accident rate of each country of the driver's origin.

Another focus of cross-cultural research is on driver aggressiveness, defined by

Lajunen et al. to be “any form of driving behavior that is intended to injure or harm other road users physically or psychologically. ” [124]. Driver aggressiveness scales [41] has been used to document differences between Serbian and Romanian drivers [175], driving anger in Spain [192], causes of driving differences of drivers in China [66], and differences in driving between urban and rural U.S. drivers [40]. Driving skill has also been posited to cause the difference between cultures. However, research (also by Özkan et al. [165]) examining that hypothesis using the Driver Skill Inventory [123] found mixed support for this hypothesis.

One issue with profiling cultural differences in driving with questionnaire-based surveys of driving is that these methods treat cultural differences as the accumulation of individual or personality differences of the people from that culture. These methods cannot easily interrogate the *social* aspects of driving culture. (These studies may also feature confounds as people from different backgrounds or cultures might have more or less self-awareness of or willingness to disclose their driving skills or behaviors [8].)

4.1.2 Capturing Driver Interaction

While driving style research focuses on the differences in the aggregation of individual behaviors, driving interaction researchers are concerned with the interactions *between* drivers as the defining characteristics of regional driving style. Sociologist Dale Dannefer, for example, mentions informal norms such as following distance, merging behavior, and right-of-way rules, but also a performance of attention or inattention [36]. Factor et al. extend this perspective, arguing that some crashes are not the result of individually risky behaviors but rather the results of “social accidents,” caused by interactions between people from different social groups interpreting and responding to situations differently [52].

Ethnographic Study

Until recently, most of the research on driving interaction was based on direct or recorded observation. Juhlin, for example, employed ethnographic techniques by observing students at a Swedish driving school, interviewing participants, recording driving sessions, and transcribing and thematically coding incidents of cooperation between road users [106]. Similar investigations have been made of social agent navigation in urban traffic [198], driver-bus interaction, [161], pedestrian-vehicle interaction [43, 191] and interactions at petrol stations [161]. Vinkhuyzen and Cefkin used ethnographic techniques to understand how AVs will engage with pedestrians, bicyclists, and other cars in a socially acceptable manner. They noted the difficulty of making observational distinctions with these methods [205].

Multi-driver Simulator Studies

Zaidel posited the possibility of formalizing the interactive model between drivers as a mathematical model that would enable the prediction of behavioral mixes in 1992, suggesting that computer and laboratory simulation would be helpful methods for beginning the research. Actual simulator studies of driver interaction are recent phenomena [217].

While many outside the automotive research domain assume that high fidelity and high immersion simulation is necessary for an ecologically valid driving response, guidance from driving simulation experts indicates that *appropriate simulator fidelity* provides the greatest fidelity for the aspects of driving under test is what is critical [128]. Driving simulators allow experimental control of conditions, reproducibility, ease of data collection, and the ability to test potentially dangerous situations in real life [38]. Even without perfect ecological validity, driving simulator studies can help researchers focus on factors or behaviors to study in follow-on research.

Driving interaction studies have primarily been made possible through multi-driver simulation platforms. Using multi-driver simulation studies to examine the interaction between drivers was first performed by Hancock and De Ridder in 2003 [86]. They placed two participants into adjacent full-vehicle simulators that share a single virtual world to understand collision avoidance behaviors. More recently, Muhlbacher et al. 2011 developed a platform to study interactions between four drivers in a platooning scenario [154]. Researchers at the Institute for Transportation Studies at the German Aerospace Center (DLR) created a Modular and Scalable Application Platform for ITS Components (MoSAIC) in 2012 to understand interactions between V2V connected vehicles and non-equipped vehicles [62, 90]. Their setup features multiple modules of high-fidelity driving simulation, such as three-display fixed-base driving simulators with a complete vehicle seat and driving interface. These researchers noted the possibility of using such a multi-driver simulator to study the effect of varying levels of drivers' experience or different cultural backgrounds or to study the influence of social psychological phenomena in traffic, such as the merging-giveaway interaction [162]. They have published studies using this setup to study cooperative lane-change maneuvers [90] and traffic-light assistance [173] Houtenbros et al. used linked fixed-base driving simulators to study whether audio-visual feedback would help participants in their interactions with other drivers; in their study, a research experimenter drove one of the vehicles [96].

Other research, particularly targeting other road users, has also used multi-participant simulation setups. A recent publication by Abdelawad et al. [2], for example, compares the aforementioned MoSAIC system, the Tokyo Virtual Liv-

Multi-driver Simulator Research	**Study focus**
Hancock and De Ridder 2003 [87]	Collision Avoidance
Muhlbacher 2011 [154]	Platooning
MoSAIC/DLR Heesen et al. 2012 [90] Friedrich et al. 2013 [62] Oeltze et al.2015 [162] Rittger et al. 2015 [173]	Cooperative driving Overtaking Platooning Traffic light assistance
Tokyo Living Lab Gajananan et al. 2013 [64]	Rubbernecking
Houtenbros et al. 2017 [96]	Audio-visual support for intersections
Feierle et al. 2020 [56]	Driver-AV interaction

Table 4.1: Multi-driver Simulator Research

ing Lab networked driving simulation (which is built on OpenStreetMap and CityEngine tools [64]), and the Driving and Bicycling Simulation Lab at Oregon State University. They mention using the setup for training drivers, for studies of truck platooning, or hybrid traffic scenario enactments. A recent collaboration between the University of Wisconsin-Madison and the University of Iowa researchers tested the feasibility of conducting driver-pedestrian simulator experiments with multiple people [108]. At the time of its original publication, *StrangeLand* was the first proof-of-concept, and we have found no publications yet describing studies designed or run on the platform.

While many multi-participant simulator systems are designed to be hybrid, each individual station tends to be quite large. This is because many interaction scenarios, such as merging or four-way intersections, require a wide field of view for each participant to see one another. (Platooning is an exception; since the main activity is ensuring your vehicle does not run into the car in front, the broad field of view is unnecessary for platooning interaction simulations.) This necessitates multi-screen or multi-projector set-ups; these scenarios cannot be run naturally using a single-screen interface.

While commercial gaming systems, such as Grand Theft Auto V, which enable multi-player interaction, have been widely available for some time, attempts to use these systems for serious driving research have been foiled or shut down by the gaming company [197]. In any case, such mods do not record critical data about the position and behaviors of each driver's car for post-analysis and study and have not been validated to produce differences in driving where we expect to see them in regional driving culture.

4.2 Original Implementation

Virtual Reality Driving Simulation Our system builds on previous multi-participant driving systems using virtual reality for networked driving simulation. The advent of networked head-mounted virtual reality platforms makes it possible for participants to have a wide field of view without having a sizeable fixed-based simulator. While driving simulation was one of the motivating uses of early virtual reality [17, 109], the use of virtual driving simulation for the experimental study of driver behavior is still relatively new [99, 194]. Virtual reality headset technology makes it easier to recreate the immersion and peripheral cues usually associated with bulkier three-screen or curved-screen driving simulation set-ups. Early research suggests driving performance is similar to that of desktop driving simulation platforms [25].

Lightweight, consumer-grade virtual reality platforms also make multi-driver interaction simulation easier to deploy in more places; this is critical to the goal of understanding cultural differences in driving. No previous system of multi-participant driving simulation using VR has been built for this purpose. The closest such system that we have learned of in our background research is a project by researchers at the University of Leeds and the Lincoln Center for Autonomous Systems in the UK. They used VR and participant tracking to have two people with VR headsets walk freely across a space and play a game of ”Sequential Chicken” with their vehicle avatars in a driving simulation environment [24]. That system illustrates the feasibility of the proposed system in this project but does not map ingrained driving interaction behaviors to virtual driving as our proposed research would.

The StrangeLand simulator uses common virtual reality hardware to make the system portable, low-cost, localizable, extensible, and accessible to more researchers. Our system builds on a body of work in the realm of multi-participant driving simulation to enable controlled experiments with common scenarios in a safe and repeatable fashion. As part of the design effort for this project, we made portable equipment set up for the driving simulation experiments, developed a multi-participant virtual reality (VR) driving simulation environment, and created and tested interactive driving scenarios. We instrumented the equipment to capture participant driving behaviors and implicit and explicit interactions. Our platform for analyzing our data streams allowed us to subjectively evaluate each driving interaction by giving researchers the ability to replay and analyze the driving interactions.

Here we describe the design of the system and experimental protocol of StrangeLand.

4.2.1 Setup

Since its original publication in 2022, the StrangeLand simulator has received significant updates to its hardware and software. In this section, we will first describe the original system and then, at the end, describe the improvements. The hardware setup for the StrangeLand simulator uses consumer-grade virtual reality (VR) gaming components.

The functional components of the system are as follows:

Laptop The simulator runs on two Alienware 15 R4 Laptops (Intel i7 8750H CPU & NVIDIA GTX 1070). Each laptop drives one VR headset.

Head Mounted Display We used the Oculus Rift CV1 VR headset for development. The hardware could be any VR headset that supports OpenVR/SteamVR.

Hand tracking To record and render the participants' hands in the virtual world, each headset has a LeapMotion hand-tracking device mounted on the front. Rendering the participants' hands in VR helps the participants to feel present in the simulation environment, and enables them to use their hands to signal with other participants. While each participant always sees the rendering of their own hands inside their vehicle, the participants can only see each other's heads and hands rendered when their vehicles are within 20 meters of one another in the virtual environment.

Drive Interface Each participant used a Logitech G29 gaming interface, with force feedback steering wheel and gas/brake foot pedals, to drive their virtual car. These control surfaces are similar to what participants are used to in everyday driving and hence are more likely to yield naturalistic driving behavior.

Network Router The computer for each driver is connected to a standard local area network connected through an ASUS RT-AC5300 router. (The system also ran successfully on other LAN routers.) Currently, the only requirement is that the IP addresses of the laptops need to be fixed.

To make it easier for researchers to transport and deploy studies to different geographical locales; we designed the system to be portable. We also designed the system to be relatively inexpensive; currently, the two-person setup for *StrangeLand* costs about $ 5000 USD.

Deployment Flexibility

We selected parts with the goal of fitting the parts for the whole system (minus the laptops) in one large suitcase (76 x 48 x 29 cm) that weighs less than current U.S. airline limits for overweight baggage. In the below Figure 4.2, one can see how the components can all be packed into one single suitcase. Important to note that the rugged parts of the setup (e.g., mounting points of the steering wheel) are facing the outside, protecting the more fragile components like the HMD. Also, the laptops must be transported as carry-ons rather than checked baggage due to current limitations on cargo transport of lithium-ion batteries.

4.2.2 VR Driving Simulation Environment

As with the hardware, the software components of *StrangeLand* were selected and designed to make it easier to deploy and replicate studies and add and extend the platform. The software is based on widely available popular software packages and is all low-cost or free.

Game engine The simulation was built in Unity 2018.4 using the now-legacy built-in networking library to synchronize the two clients [201]. This enables participants to see other drivers in their virtual vehicles and note their head orienta-

Figure 4.2: The complete *StrangeLand* simulator hardware in one suitcase (with packing foam removed). The updated version requires even less space and hardware. See Figure 4.3

tion or exchange hand gestures.

Vehicle model We used a prefabricated vehicle model and logic from the GENIVI [68] driving simulation platform. We extended the model to include a car interior and interaction logic for the steering wheel, horn, and indicator lights.

Head mounted display interface The main VR interface used in Unity is the OpenVR library [203] connecting to SteamVR [202].

Hand tracking The Orion software pack from Leap Motion for the hand tracking in combination with VR. [200]

Environment elements We developed the road track in which the different scenarios took place was modeled using openly accessible textures. Buildings were placed at the corners of the intersection to ensure participants could not see the entire track without approaching the intersection.

4.2.3 Improvements in the Version from 2023

In 2022 and 2023 the code was refactored and updated. In this section, we will describe the changes to the system and what it means for the simulator.

Headset Technology

Since the original release of *StrangeLand*, headset technology has improved significantly. The previous version used the Oculus CV1, a headset that is no longer available and that used an external computer to render to the display. New headsets now incorporate the computer that renders the display directly in the headset. The new hardware also performs inside-out tracking, which includes hand tracking. The new implementation uses these new hardware and features.

Since the rendering is done on the headset, the setup only requires one laptop now acting as a server, the headset connects as the client.

Hand tracking used to be done by the *LeapMotion* controller can now be handled internally by the headset without any additional hardware, further simplifying the deployment of the system. the main drawback of using a standalone headset is the limited CPU/GPU performance. Some materials and models needed to be optimized and simplified to make sure the headset could render at a high frame rate throughout the study.

Network Code

In 2019 Unity deprecated the network code used in the original *StrangeLand* implementation. after a few years of no officially available net code, Unity released *NetCode for GameObjects* in 2022[2]. This implied that substantial parts of the system

[2] The official 1.0.0 release was on 2022-06-27 `https://github.com/Unity-Technologies/com.unity.netcode.gameobjects/releases`

Figure 4.3: The updated version of *StrangeLand* requires even less space. All components fit easily into one rugged travel case. *Credit: Hauke Sandhaus*

needed to be rewritten. While there are many detailed changes, the main topological changes in the internal system are as follows:

1. All simulation now happens on the server; this includes all participant input capture (e.g., steering wheels), data recording, and rigid-body/vehicle simulation.

2. The headset needs to follow the simulated data and "mostly" renders the environment. That implies that more data needs to be shared between the headset and the server, such as indicators and the use of the car horn.

3. The questionnaire data handling needed to be completely reworked to allow for editing of the questions without recompilation. Unrelated to the network changes, the new implementation of the questionnaire screen also included showing pictures during the questionnaire to contextualize the questions for

the participant. These could either be predefined pictures in unities *Resource* folder or a screenshot taken during the study.[3]

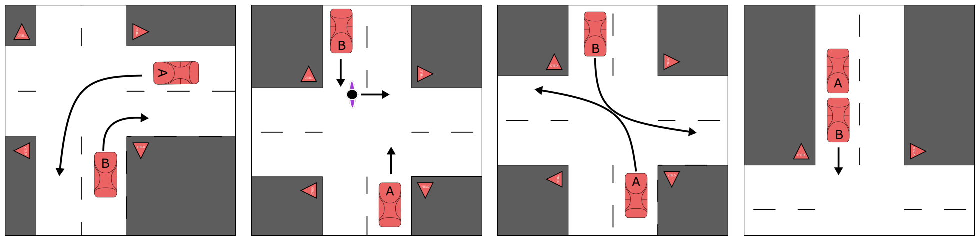
The practical implications of these changes are that less hardware is used; the simulator can be deployed with just two headsets, steering wheels, and a mid-tier gaming laptop (see Figure 4.4). More implications of this re-implementation can be discussed below (see section 4.5.1).

Figure 4.4: Setup of the re-implemented *StrangeLand* from 2023.
The Screen in the back, while helpful for demonstrations, is not required to run the simulator.

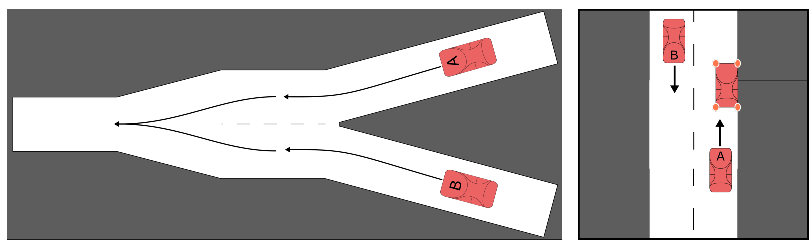
4.2.4 Interactive Driving Scenarios

To capture driver interaction, we developed traffic scenarios that required drivers to negotiate with one another to complete their driving tasks. For example, we designed an intersection with a four-way stop scenario where multiple drivers

[3]If a screenshot is taken, the image is serialized and sent back to the server and amended to the data storage.



(a) S:1 (S:2) Orth. A Right (b) S:3 Opposite Pedestrian (c) S:4 Opposite Left (d) S:7 (S:8) A Behind B



(e) S:5 (S:6) Merging A Right (f) S:9 (S:10) A Blocked Lane

Figure 4.5: Showing all the types of scenarios developed for the study. Scenarios that are the exact mirror of another scenario are not shown. Instead, the mirrored version is referenced in parentheses in the image caption.

arrive at approximately the same time. Because it is difficult to decisively determine who arrived at the intersection first, drivers need to observe each other and negotiate who will go first to avoid colliding. These scenarios were intended to elicit routine interaction responses that drivers use every day. We manually selected the driving scenario, counterbalancing the order of scenarios across participants after iterative designing and optimizing the scenarios through pilot studies.

We tried to account for the inconsistency of signage and road standards across different locations to enable cross-cultural studies. For example, yield signs have a consistent meaning across cultures [211] (although with slightly different standards about height and placement [81]), so we tried to use more yield signs than stop signs, which have greater cross-cultural variance.

We designed and tested several scenarios to ensure that drivers were clear on their driving goals but not on the right of way concerning the other driver. We also designed the scenarios to be counterbalanced so that both participants in the study had a roughly equivalent experience. So, for example, if one participant turns left and the other turns right, we include the reverse scenario. Here is the resulting set:

S:1 - *Four-way Intersection:* Car A and B are orthogonal to each other at a four-way intersection (A begins on the right). Car A must turn left while Car B is instructed to go right. The two cars are turning towards each other.

S:2 is the counterbalanced of four-way intersection scenario **S:1**.

S:3 - *Intersection with Pedestrian:* Car A and Car B are instructed to go straight at opposite sides of an intersection; as the cars approach, one pedestrian will start to walk across the street. The pedestrian has, by design, an ambiguous starting time, as they only begin moving as either car A or B approaches.

S:4 - *Opposing Left Turns:* Both Car A and Car B appear at opposite sides of an intersection, and both receive instructions to turn left [174].

S:5 - *Merging:* Car A (right) and Car B (left) are merging onto the highway from their own respective roads. In most merging situations, it is clear who has the right of way because one car is merging onto the road of another car. However, in this scenario, both roads merge into the same road giving neither right of way.

S:6 is the counterbalanced merging scenario **S:5**.

S:7 - *Overtaking:* Car A is driving behind Car B. Car B is instructed to "stop", while Car A is instructed to "Hurry up." Car A must decide to overtake Car B. Driver of Car B is unaware of the instructions to the driver of Car A, leading to uncertainty about their action. It is important to see when, how, and if they decide to overtake Car A.

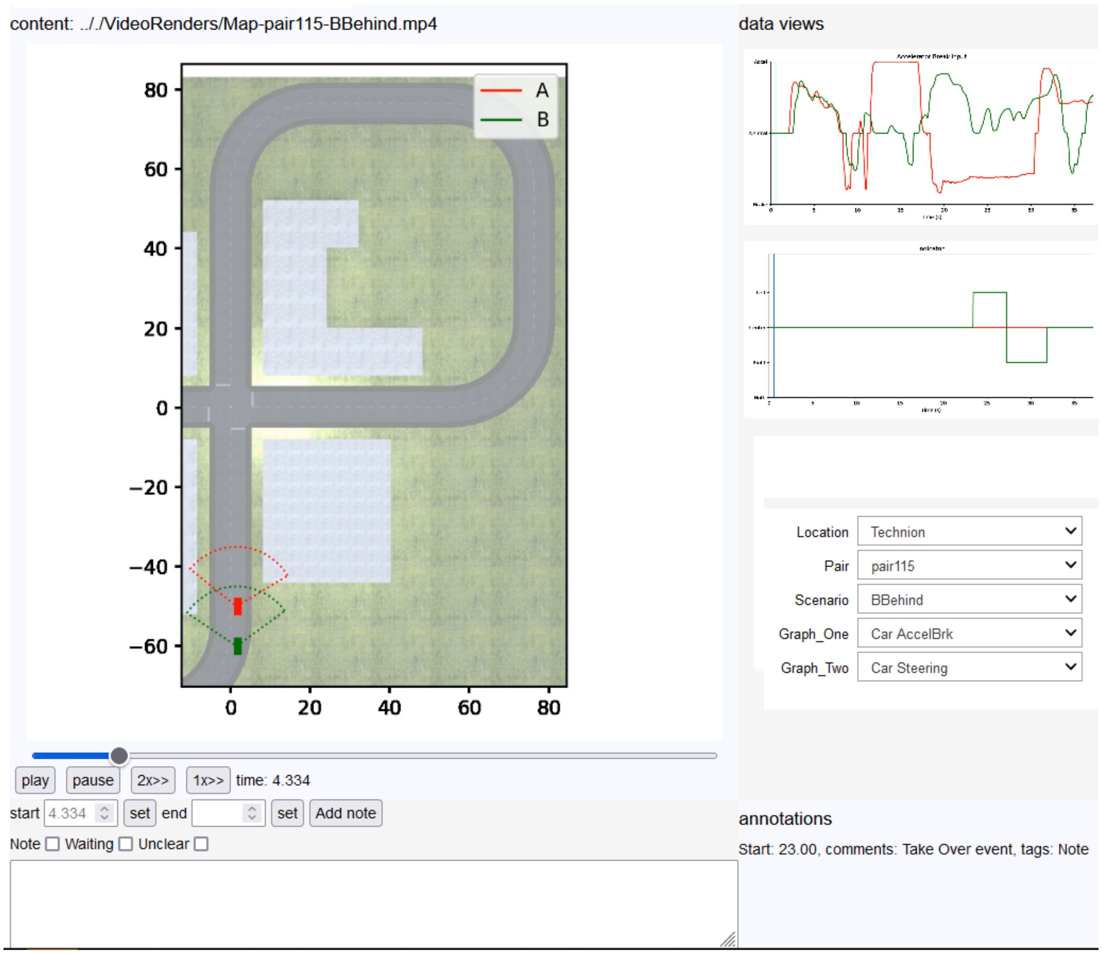


Figure 4.6: Screenshot of the behavioral analysis tool. At the top left and top right, we see the map and graphs videos showing the behaviors as they took place. At the bottom, there is space for writing (bottom left) and reading (bottom right) annotation for this particular participant pair and scenario.

S:8 is the counterbalanced overtaking scenario **S:7**.

S:9 - *Blocked Lane:* In front of Car A, there is a parked car with hazard lights blocking the lane while Car B is approaching the oncoming lane. As a result, the driver of Car A has to decide whether or not to wait for Car B while the driver of Car B can choose to stop and let Car A pass.

S:10 is the counterbalanced block lane scenario **S:9**.

4.2.5 Behavioral Analysis Support

Because our driving simulator intends to capture a range of interactive behaviors which we expect to differ as a function of the drivers' cultural norms, one key aspect of our simulator design is that it needs to support the observation and analysis of the communicative actions of drivers. Typical driving simulation studies often measure performance or driving behavior in response to pre-defined stimulus events which occur in a controlled environment. Our simulator also contains a controlled environment but, in other ways, is more like a naturalistic study of group interaction; researchers observe how the interaction emerges between the participant under different controlled circumstances.

To enable a qualitative analysis of driver interaction, we developed an interactive behavioral analysis tool on a Jupyter notebook (Figure 4.6). This tool allows us to reconstruct and analyze interactions from multiple viewpoints reconstructed from the generated data. The notebook includes a map view, speed, and indicator line graphs, in addition to synchronous video data from the VR world. These multiple viewpoints enable qualitative as well as quantitative analysis of the interactions. An example of the output from the interface can be seen in Figure 4.7. In this figure, the accelerator brake input and steering input is recorded. In the beginning, one can see that the steering input is at the center when the vehicle accelerates.

Map View We generated videos with a map view of the car based on the simulator data (speed, location, head orientation). This top-down view allows us to intuitively examine the traffic scenario and discover behaviors from the participants, e.g., how some participants continually creep into an intersection.

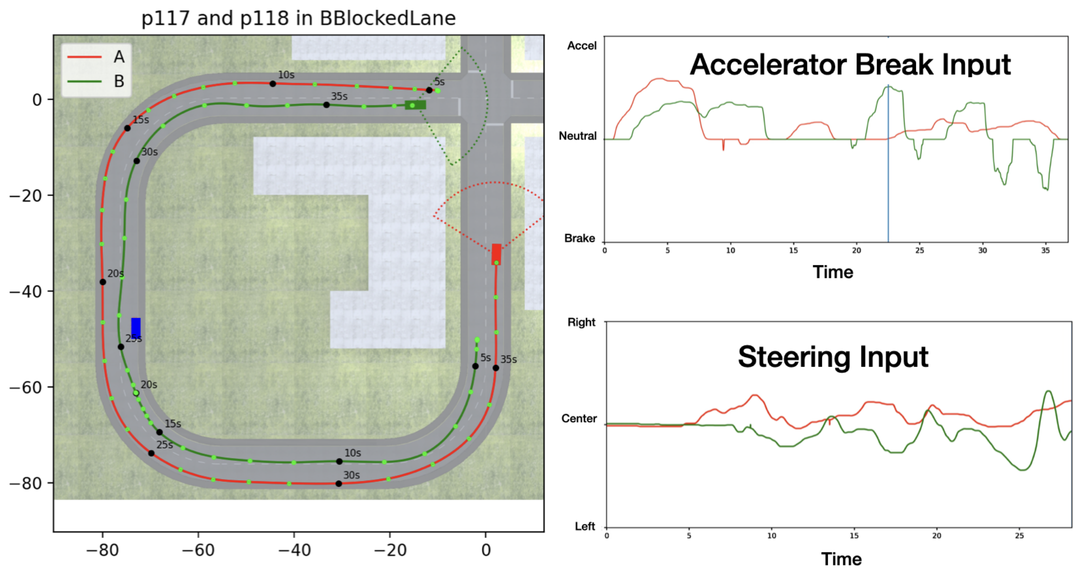


Figure 4.7: Top-down view of the virtual environment. This shows the path of two participants in the scenario, Blocked Lane. The green tick marks indicate the position over time. The y and x-axes are measured in meters with the intersection at 0/0.

Steering, Speed, Paddle and Indicator View Additionally, we generated animated graphs to analyze the measures and played them back in conjunction with the map view. These graphs give a more detailed look at the participant's responses. E.g., it is easily visible when and how strongly someone slowed down in reaction to an incident or event.

Video Data In addition to the generated data view, we can playback the synchronous video data from the GoPro and the respective laptops' screen recordings. This video data allows for a subjective first-person evaluation of the "normalcy" of the interactions.

In addition, we use an **in-simulation questionnaire** to assess the situation awareness of each participant, structured using three levels (perception, comprehension, and projection), known as the SAGAT method [50]. We used this method to avoid taking participants out of the virtual world many times throughout the

Figure 4.8: Post-interaction questionnaire internationalized in (a) English and (b) Hebrew.

study. An example is shown in Figure 4.8. Within the VR simulator, both participants are prompted to answer questions that appear on a translucent screen in-world after each scenario. The first question of the questionnaire would begin consistently across scenarios, asking about certainty (i.e., "I clearly understood the intent of the other driver(s)").

In total, seven different questionnaire sets were asked through this VR method. These included fact questions (i.e., "At the intersection, who moved first?") and then understanding the facts (i.e., "Why did you move first?"). While this structure of *fact & understanding* remained for the other question sets, the topic differed (i.e., turn signals, stopping, who moved first, false starts, overtaking, eye contact, cutting off). We chose to ask about a particular topic based on the scenario. For example, in the four-way stop scenario, the participants would be asked about turn signals, who moved first, false starts, and eye contact but not about overtaking, as this did not happen in this scenario. (A complete list of all possible questions can be found in Appendix A.3.1)

In the 2022-2023 updated version of *StrangeLand*, we also implemented an additional recording tool that fully reconstructs the scenario in 3D. The tool is rather versatile and can be used to record all kinds of interactions happening in VR. It is called **ReRun** and the subject of Chapter 5.

4.2.6 Instrumentation of Behavior and Interaction

By logging data about the participants, their behavior, and the state of the virtual world throughout the interaction scenarios, we can collect key measures that we believe are instrumental to understanding driver behaviors and their interactions with each other. Many of these measures were informed by SAE's J2944 Operational definitions of driving performance measures and statistics [163]. These quantitative measures are particularly important as a secondary step to verify findings from qualitative video plot analysis findings.

We describe a non-exhaustive list of possible measures that could be analyzed out of the given data below:

Hand pose: we collect the hand pose and articulation over time through the Leap Motion or build-in hand tracker. We can tell if the participants have their hands on the steering wheel, whether they are steering or waving to someone.

Head orientation: Through the Oculus Rift or Quest 2, we can collect the position of the head relative to the world. From this data, we can tell if car B, in the dyad, is in the field of view of car A. SAE J2944 does not have any recommendations regarding head orientation. However, they guide the need to measure where drivers are looking, particularly for lane change tasks. [129, 163] It is possible that in longer scenarios than what we tested here, researchers could also use this measure to infer distraction and fatigue. We can use this in our interaction scenarios to see if drivers are in each other's field of view at different points in their interaction.

Steering Direction: Through the steering wheel, we collect the rotation information of the steering wheel's position. This, combined with event logs from the simulation environment, allow us to measure steering reaction time, movement time, response time, and steering reversals. The steering input is recorded directly from the steering wheel as a floating point number (-1.0 Left; 0.0 center; 1.0 right).

Pedal Input: Through the gaming interface, we can measure the participant's input to the accelerator and brake. This can be used with simulation environment event logs to infer the accelerator and brake response times.

Car Position, Velocity & Acceleration: In the simulator, we can determine the position, speed, and heading for each car at each moment in time. Additionally, we also store the car's velocity. This allows us to measure lane position, lane and roadway departures, as well as lane changes.

Wait time: Through timestamped data and car positions, we can analyze the wait time until the participants move in ambiguous parts of the scenarios.

Entropy/Energy: By summing the cumulative difference in longitudinal or lateral input, we can obtain an "energy" measurement that corresponds to steering reversals or excessive changes in speed. For lateral input, this corresponds with the measure of Steering Entropy used in SAE J2944 [157, 163].

4.3 Proof of Concept Testing

To ensure that our system can produce meaningful measurements that enable comparing driving behaviors across cultures, we performed a proof-of-concept test. The aim is to describe the design of a system that allows the capture of important cross-cultural differences in driving, the test aims to establish that the system we built is functional, deployable, reliably captures data, and enables reconstruction of interactive behaviors. Full-scale study deployment and results featuring

Confidence Questionnaire	Answer Options				
I felt the other driver drove well.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I clearly understood the intent of the other driver.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I felt confident about my own actions.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

Table 4.2: A table showing the questions and answers for the confidence questionnaire.

claims about differences in cultural driving behaviors will be attempted and detailed in future work.

During the initial development, the researchers ran an iterative design experiment (N = 10) in Haifa, Israel, at the Technion. Based on the input from these studies, we further developed the simulator, particularly the scenarios. Some of the improvements we made during these studies were e.g. adding virtual mirrors to the car, giving the participant a horn, and adjusting for both short and tall participants. After these experiments, we ran more tests at the Technion and Cornell Tech to further test and verify the system's stable operation and generate data to develop and test the data analysis pipeline that could be developed and tested.

4.3.1 Original Study

To ensure the effectiveness and reliability of the initial version of *StrangeLand*, we conducted a proof of concept study. This study aimed to validate the simulator's ability to produce meaningful data that can uncover various modes of interaction. By running this preliminary investigation, we aimed to establish the simulator's capabilities and its potential to provide valuable insights into driver-driver interactions.

Participants

For the Proof-of-concept test at Cornell Tech, we recruited using convenience sampling and had N = 26 participants (18 male, 8 female) between 21 and 41 years old, with an average age, M = 26.5, SD = 4.45. 24 participants learned to drive in the United States; one participant learned how to drive in Costa Rica, and one in India. Of the 26 participants, 3 got motion sickness from driving in the virtual reality environment. Keeping the headset on during the questionnaire made it so that participants kept HMD on between vignettes, did less context switching, and shortened overall time in the experiment. Participants had between 0 and 22 years of driving experience (M = 9.3, SD = 4.25). When asked how many times they drove a week in the last year, the answers varied widely, but about half stated that they had primarily driven three times per week or only when they were in the city. We also asked where else people have driven for more than one year, outside of the United States: one participant said Canada, and one said Israel. In both places, we used a between-subjects study design.

At the proof-of-concept test at the Technion, we recruited using convenience sampling we had N = 52 (31 male, 21 female) between the ages of 21 and 33 (one participant was 52). 27 participants knew the other participant in the experiment, and 25 participants did not know the other participant. All of the participants learned how to drive in Israel. Out of the 52 participants, 4 got motion sick. Participants had between 0 and 30 years of driving experience (M = 6.86, SD = 4.47). Most participants stated that they drive more than five days a week. We asked participants where else they have driven for more than one year, if at all, outside of Israel; 5 participants answered yes; one participant said Romania, one Mexico, one Germany, one United States, and one Ibiza.

Although some participants in the U.S. study were not originally from the U.S., for the proof-of-concept test and to have a comparable number of participants across the U.S. and Israel, we have decided to include all pairs in the analysis. This inclusion may seem less controlled but, in fact, maybe ecologically valid, as the U.S. site features more tourists and international visitors, and so greater cultural variance even within the geographical locale is the norm.

Procedure

When participants arrive at the study room, they are led through the informed consent process and are told about the remedies available to limit nausea, like ginger candies and wet towelettes for the forehead. Next, we start the data recording on the GoPro. Participants each completed a demographic survey. Next, the participants are informed about operating the system: pedals, steering wheel, horn, VR headset, GPS, hazard lights, and turn signals. They then are told how to answer the questionnaire in VR, using eye gaze to rest on their multiple-choice answer. They are instructed not to speak to each other verbally but only to communicate in the virtual world. They are told that they may stop the experiment at any time if they feel uncomfortable. [190]

Next, the participants put on the VR headset, put their hands on the steering wheel, see their hands on the steering wheel, and their feet near the pedals. We then calibrate the VR tracking system, aligning the virtual world to the physical world, using the steering wheel and tracked hands as a reference point. Once the system is calibrated, we tell the participants to drive around an empty course alone to get familiar with driving in the virtual reality world.

We ask if the participants are ready before beginning the interaction traffic scenarios. When they are ready, we manually select the driving scenario, counterbalancing the order of scenarios across participants. Next, the participants drive in a given scenario and then answer questions in-world about what had just happened. After five of the scenarios, we ask the participants to take a break, to prevent nausea. Following that, the participants continue the same process for five more scenarios. Finally, to conclude the study, the participants are asked to take off their headsets, are given compensation, debriefed on the experiment, and be thanked.

In Israel, the average time of each scenario (from the start driving until the end trigger) was $M = 38.58$ seconds, with a standard deviation of $SD = 17.69$ seconds. In the U.S., the average time of each scenario was $M = 34.57$ seconds, with a standard deviation of $SD = 24.08$ seconds. The average practice times in Israel were $M = 110.18$ seconds, $SD = 30.95$ seconds, while in the U.S., the average practice time was $M = 64.22$ seconds, and $SD = 26.52$ seconds.

4.3.2 Findings

Part of our proof-of-concept study deployment was intended as a proof-of-concept test to see whether and how well the *StrangeLand* platform achieved the technical requirements needed for cross-cultural driving interaction research. The system needed to be readily deployed in various locations, present the same context and scenario across different areas to elicit behavioral differences, and support naturalistic interactive behaviors between drivers. We discuss our assessment of these criteria here and then further discuss the interactions and driving behavior captured by the system.

Deployability

Because we intend for cross-cultural simulation studies to occur in various locations, the *StrangeLand* simulator must be transportable and deployable in various lab, office, or conference room settings. This study's two locations were intended in part to show the practicability of the system for transport. We will also mention anecdotally that the system was relocated several times and in three different countries during development. Setup time for the simulation equipment can be well under an hour if chairs, tables, and power outlets are available. The equipment is also based on commercially available gaming and entertainment hardware, so the bulkiest parts of the *StrangeLand* setup, the steering wheel, can also be purchased at each study site for $\leq \$1000USD$.

Controlling Scenarios Across Locations

For the proof-of-concept study, we were able to have participants in our study drive in exactly the same scenarios in both of our study locations. This was desirable in the proof-of-concept study because we wanted to verify that we could elicit differences in interaction and behavioral measures across two sites and avoid the confounds that would occur if there were any differences in the virtual surroundings.

The system and scenario development required numerous iterations to solidify the overall protocol. A number of study design dilemmas emerged during scenario development, such as the fact that four-way stops, for example, which are prevalent in un-signalized intersections in the US, were not at all typical in Israel, where traffic circles are common. There is also some tension between localizing the study environment to be typical and familiar to the drivers and keeping the study environment a little more abstract. For example, SUVs are more typical in the US, and compact cars are more typical in Israel. Ultimately, we decided to use the same buildings, cars, and environments in both our study locations for experimental control; if we had varied the environment, however superficially, we would have had to perform validation experiments to see, for example, people would be more likely to yield for one type of car or another.

One important thing to note is that the design of the *StrangeLand* system makes it possible for other researchers to test the effect of such variations. *StrangeLand* is implemented on the widely accessible Unity game engine. Since many other simulators use this engine base, their 3D graphic assets and software libraries can be reused with *StrangeLand*; this enables flexibility in the scenarios and extends *StrangeLand* use for other environments. This flexibility and reuse is an explicit goal for *StrangeLand*, especially, after its redevelopment, in which many code elements were rewritten to make them more flexible in how they can be used. We did not employ any proprietary graphic assets in *StrangeLand* that we would not be able to distribute, although employing such graphic assets could improve the visual appeal of the simulations. As designed, it would be easy for researchers in other locations to add scenarios or skin the cars, buildings, or signs to *StrangeLand* as they deem appropriate for their studies. By making the *StrangeLand* system open-source and sharing our study datasets, we make it easier for researchers in other locations to set up comparison studies and directly compare their results to ours.

Immersion and VR Performance

At a high level, our goal is for participants to feel immersed enough in the simulated environment to interact naturally with other drivers. Our goal is to elicit the natural differences in driving that people practice. Part of this, we felt, was that participants needed to handle the alignment between their physical actions and that of their virtual avatar and to be able to interact with the other participant as they would with another person in the real world.

During the studies, all participants had no problem operating the virtual vehicle or associating with the virtual representation of their hands. Qualitatively, we observed numerous episodes where participants responded to the gaze and gestures of the other participant in ways that suggest that the *StrangeLand* platform supported their naturalistic interactive behaviors. Quite a lot of gesturing occurred (see the participant on the left in the Figure 4.1, for example). Anecdotally, gestures from one participant caused return gestures from the other participant. This is significant in part because *StrangeLand* is, to our knowledge, the first driving simulator environment which tracks and renders the hand gestures of participants, and so is the first system to be able to capture this type of interactive behavior.

For a more quantitative verification of the function and immersivity of the VR simulation environment, we examined the frame rate, external perceived motion-to-photon delay and the network delay by comparing time differences from participant study runs during the development and after the proof-of-concept study.

For VR applications, the frame rate should be greater than 60 f/s to create an immersive experience [63]. Data analysis from the proof-of-concept studies showed a median frame rate of 90.9 f/s with a standard deviation of 6.0 f/s. This means that most frames (> 95%) were rendered within 70 to 90 frames per second. The researchers' subjective experience supports this finding during development and testing, during which no noticeable stuttering occurred. Adhering to these standards in the updated version was more challenging, as the system needed to render on the limited on-device hardware.

To ensure that participants did not encounter extended periods of stutter, we computed a 1d convolution over a window of 4 frames, which reduced the standard deviation by about half to 3.1 f/s. This finding shows that often a slower

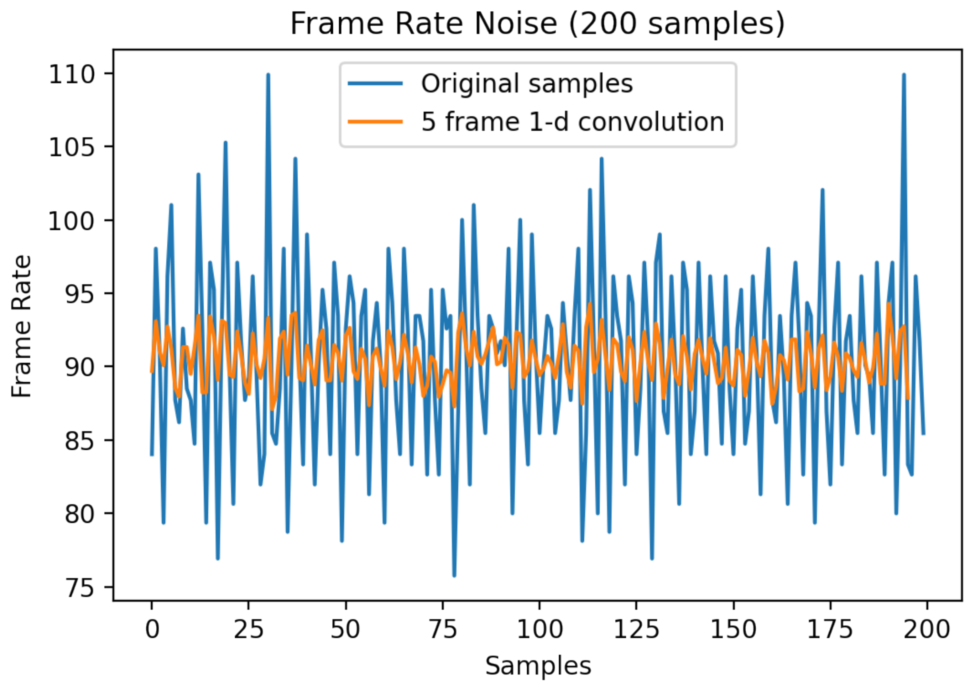


Figure 4.9: Showing how the f/s vary significantly between each frame and how a small 5-frame 1-d convolution shows the continues f/s.

frame was preceded by faster frames and that it was exceedingly unlikely ($<$ 0.001%) for any participant to experience a frame rate of less than 80 f/s for more than 5 frames in a row.

To verify both motion-to-photon delay and network delay, we used a GoPro action camera (set at a 29.97f/s setting) to take a video recording with both participants in the study setting and their respective virtual views on a laptop screen. Looking frame by frame at the head and hand motion of the participants, we could not measure a single frame difference between head motion and the rendered frame appearing on the laptop screen. The headsets were operating in direct mode, which has less delay than the laptop screen used to measure the delay. The video was recorded at 29.97f/s; this sets the upper bound for the frame delay to be 33 ms.

The same video source was used to probe the network delay between the two participants. In particular, the GPS display and the questionnaire screens are network-synchronized events that use the same network bus used for the transform and hand data. Therefore, the network delay should be consistent. As with the motion-to-photon delay, the events appear in the same frame; this implies a network latency equal to or less than 33 ms. Studies on networked multiplayer video games, in particular racing games, set the acceptable latency range at 50 ms [28, 166].

Figure 4.10: We used video recordings with the participants' screen captures in view to validate both the motion-to-photon delay and the network synchronization delay.

Interaction in the Designed Scenarios

Using *StrangeLand*'s analysis platform, we can replay scenarios to view how the cars interacted from an overhead perspective. For example, we can see how two vehicles slow down as they approach each other. Observing this behavior helps us verify that the two drivers are aware of each other.

Since we are interested in studying driving interaction, we looked to see which scenarios seemed to generate the most driver-driver interaction. By observing and interpreting the data from our analysis platform, we found that merging scenarios (S:5, S:6) produced the least amount of interaction. It seemed that one participant was not waiting or depending on another participant. In contrast, in the overtaking scenario (S:7, S:8), **participant A** was, by default, dependent on **participant B**, and hence waiting, and monitoring behaviors could be observed.

The analysis platform also allowed us to find scenarios that need to be re-designed to elicit ambiguous situations. One way to achieve this is by timing participants' arrival at a certain point such that the right-of-way becomes ambiguous.

Waiting on the other car We use car position as the reference to all other measures. The Figure 4.7 shows a graph from a pair of participants in a basic blocked lane scenario. In this case, participant B's lane is blocked by a parked car. The plot shows the participant slowing down; the tick marks, which are made once per second, become closer as the car is now slowly rolling towards the stopped vehicle; after Participant, A passes by, we see the subsequent left turn of Participant B.

Comfort and Confidence in Interaction We observed that the speed with which a participant drives is an important measure in understanding their driving behavior. It is closely linked to the participants' comfort driving at a certain speed given a certain traffic scenario. The Figure 4.11 shows the average speed in meters per second for the Opposing Left Turn scenario and their respective 95% confidence intervals on the second pair of bars. We can see a difference in average speed between the Haifa, Israel, and New York, New York participants. Israelis appear to

Variable Name	Variable	Data Type
Action State	DRIVE	string
Car AccelBrk	0.2744	float32
Car PositionX	-31.0257	float32
Car PositionY	0.0956	float32
Car PositionZ	-29.2038	float32
Car RotationX	359.9413	float32
Car RotationY	18.1327	float32
Car RotationZ	0.0018	float32
Car Steering	0.0164	float32
Car VelocityX	2.5615	float32
Car VelocityY	-0.0002	float32
Car VelocityZ	7.7710	float32
Event log	“ ”	string
Frame Number	568646	int32
Game Time	6060.32	float32
Head PositionX	-31.3509	float32
Head PositionY	1.1624	float32
Head PositionZ	-29.091	float32
Head RotationX	3.9596	float32
Head RotationY	22.47	float32
Head RotationZ	2.0991	float32
Left Indicator	OFF	string
Left Vectorhand	APg...	Base64 string
Real Time	1564388496.3204	float64
Right Indicator	OFF	string
Right Vectorhand	ARM...	Base64 string
Time Scale	1.0	float32

Table 4.3: This example data frame shows the generated data from the simulator. *Left* and *Right Vector-hand* variables could be decoded into individual bone positions; here, they are recorded in a compressed format to safe bandwidth.

have driven slower in that specific scenario.

Wait Time Another related measure is the wait time, which is the time between coming to a stop and resolving an ambiguous situation (i.e., the time they waited at the intersection). This measure can be found in Figure 4.11 for the Intersection

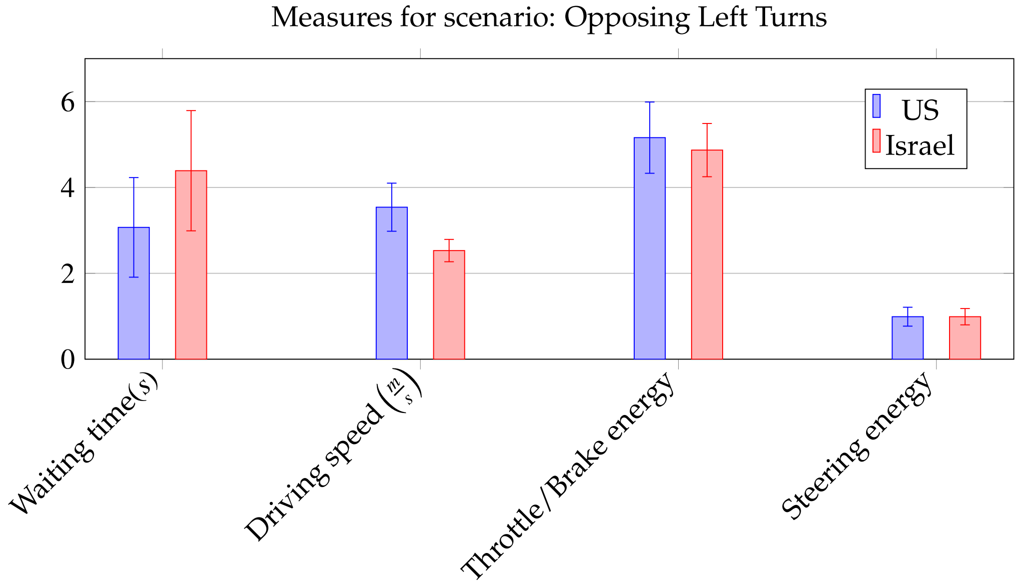


Figure 4.11: The comparison of waiting time, driving speed, throttle/brake energy, and steering energy measures for one example scenario illustrate that the system generates consistent data. Error bars show 95% confidence intervals.

with Pedestrian Scenario. The first two bars show the average wait time in seconds and their respective 95% confidence intervals. As with the speed parameter, there appears to be a difference between Israeli and US drivers, with Israeli drivers waiting longer.

Erraticness of driving Many of the metrics described in SAE J2944 could be computed based on the available data. However, we calculated a simple cumulative difference for the input parameters, longitudinal (speed) and lateral (steering) control for the example data. This basic "energy" measurement corresponds to the steering and paddle input change. This data is exemplified in the last two pairs of bar graphs, in Figure 4.11 indicating that Israel and the US participants were comparable in how they felt they needed to provide input into the simulator. This, by it itself, is a positive finding mainly because none of the participants have to over-correct their vehicle input.

Questionnaire Evaluation

Results showed that both populations were certain about their driving and the other driver's driving. Results from the 5 points Likert questionnaire show that overall positive responses were more common for all three questions on certainty, and the most common answer was "Agree." In the United States, for the question, "I felt the other driver drove well," the average response was M = 3.91, SD = 1.09. In the question "I clearly understood the intent of the other driver," the average response was M = 3.85, SD = 1.09. Lastly, in the question, "I felt confident about my own actions," the average response was M = 3.97, SD = 0.4.

In Israel, for the question "I felt the other driver drove well," the average response was M = 3.96, SD = 1.14; for the question "I clearly understood the intent of the other driver," the average response was M = 3.89, SD = 1.14, and for the question "I felt confident about my own actions" the average response was M = 3.75, SD = 0.43.

The two groups had statistically insignificant differences in their answers for the statements "I clearly understood the intent of the other driver" and "I felt the other driver drove well" with two-tailed t-test p-values of 0.187 and 0.1197, respectively. Participants from the United States agreed slightly more with the statement "I felt confident about my own actions" with a significant p-value of 0.0015.

Our situational awareness assessment asked participants what occurred in their interactions with the other driver. It allowed us to analyze the participants' awareness of their driving styles and compare their actual actions in the simulator with their alleged actions as recalled in the questionnaire. While participants in the

United States accurately recalled their own and their partners' turn signal use about 80% of the time, participants in Israel recalled their own turn signal use more accurately than the turn signals of their partners. This suggests that Israeli drivers may pay less attention to fellow drivers' turn signals.

Simulator Sickness

Both driving simulators and virtual reality experiences can cause nausea and simulator sickness. At the Technion, the Simulator Sickness Questionnaire data was collected (N = 56) pre-and post-experiment to evaluate the participants' sickness likelihood and incidence. Out of the 56 participants, four reported nausea. Overall, these results suggest that the severity of self-reported simulator sickness with the *StrangeLand* setup was low.

4.4 Discussion

Our long-term goal in creating the *StrangeLand* system is to capture cross-cultural driving differences in a manner that would allow a computer-controlled car to recognize and respond to local driving norms. As the first step towards this, we can use *StrangeLand* to elicit naturalistic driving interaction between people in different locations in ways that enable researchers to reconstruct and analyze what transpired between participants and find promising evidence of regional differences in driving culture. While this builds on prior work in multi-driver simulation systems, such as [86, 154], our platform is game-changing because it is built on lightweight, portable consumer-grade equipment using open software. This fact makes the system suitable for deployment in multi-location studies in a manner that previous systems had not achieved; this is why none of the previous systems had been used for the purpose of profiling cultural differences. The difference in cost of these platforms is one or two orders of magnitude. The low-cost and consumer-product architecture makes it possible for other researchers in other locales to build their own version of our system and run comparison studies replicating our methods with their local population.

Scenarios

The design of interactive systems that can respond to culturally-specific driving interactions requires an autonomous system to recognize interactive bids and maneuvers by people and respond appropriately. This approach of using simulation environments to elicit naturalistic interactive behaviors can also help develop other interactive products. These could be conversational agents where being savvy about local norms could improve the product. In staging the scenario, our work enables the first step towards designing future interactions; it collects data about how people in different locales currently negotiate these scenarios, giving us information about what exchanges lead to better or worse outcomes. The use of the virtual environment to collect this data reduces the effort that is needed to recognize scenarios and to control for conflating factors when trying to understand interaction in the wild, for example, as Domeyer et al. has done using data from MIT's Advanced Vehicle Technology data-set [47, 61].

As mentioned previously, the system and scenario development required numerous iterations to solidify the overall protocol; the scenarios we developed yield meaningful differences in driver behavior. Of course, these scenarios are not exhaustive; however, when compared to the proposed driving interaction framework by Markkula et al. (which was published after our system was developed and being piloted), we are pleased that our scenarios cover all of the driver-driver interactions except that where two vehicles are vying for the same parking space [136].

One of our goals in future work is to add scenarios and measures to establish a more comprehensive data-driven model of cultural driving styles. Superficially, the platform can be ”skinned” to adapt signs, buildings, and vehicles to match local regulations, regional architectural styles, and typical traffic makeup.

Cultural Driving Styles

While we do not intend these initial implementations and tests to be used for broad claims about cultural differences, we believe the findings suggest “construct validity” [34] for cultural driving differences. The early test found statistically significant cross-cultural differences in driving between the U.S. and Israeli drivers in their average recall of turn signals. Additionally, we found a significant difference in average speed between the two countries while having similar input rates.

This is a positive indication that a more extensive and controlled study with a deeper analysis of complex interaction patterns would unearth other driving differences. Furthermore, this lays the ground for future research which can profile regional differences in driving culture, which are essential for drivers and autonomous systems to adapt to.

Methodological Issues

In running the proof-of-concept tests, we noticed some issues that we think need to be addressed before the system and protocol can be used for research. For example, we noticed that, sometimes, participants were communicating verbally instead of through the simulator despite our instructions. While it is common for drivers to communicate with passengers within the vehicle, this isn’t the case between cars. This could affect the verisimilitude of the simulator. Verbal communication could also obviate the need to communicate through gesture or vehicle motion as people would in regular traffic. Therefore, we plan to amend the protocol to start the study by assessing whether the participants know each other. In addition, we will physically separate the participants to prevent cross-talk.

During the study, we found that participants did not always start when told to. This would throw off the timing of our designed scenarios and cause misses where we intended to have interactions. We are looking into programmatic and simulator-based solutions that could adjust vehicle speeds so that participants experience arrival at the critical event simultaneously.

Finally, the counterbalancing of the scenarios caused some scenarios to be experienced twice from both sides. This potentially could have made the second scenario more predictable, causing learning and interaction effects. In the future, we plan to create a visual distinction between similar scenarios by designing trivial scenery differently. By doing so, we hope drivers are less likely to recognize that they are in a mirror scenario from one they experienced previously.

Features for Studying Interaction

StrangeLand is not the first multi-user simulator. However, it was designed with a direct focus on the implicit interaction that happens between drivers as they encounter each other on the road. These features and their combination is particularly important as it allows for scenarios and findings not covered by prior systems.

Hand Tracking When deploying a VR-based simulator, hand tracking is always crucial as it gives the participants the sense of place necessary to grab and halt the physical steering wheel. This capability to share the tracked hand data with the other networked participants allows for hand gesture communication. Additionally, the head pose is also shared between participants such that another driver can see where the driver is looking.

Field-of-view By using VR headsets, there is no technical limit to the field-of-view a participant can achieve by turning their head. Many of the existing multi-participant simulators feature a three-screen setup that only covers a portion of the participant's field of view. While for many driving scenarios, this should not be concerning, it can become a limiting factor for urban driving with intersections and interactions happening at 90° or higher relative to the participant's orientation. *StrangeLand* allows the participant to look around and gain situational awareness similarly to how they would in a real car, allowing for interactions with road users coming from any orientation. This is further aided by the simulated mirrors.

Additional features for interaction have also been implemented; these, however, can also be found in other simulators.

Limitations

One key limitation of this work is that the studies in both locations were run by university students. We believe that this is appropriate for proof-of-concept testing: if you cannot get statistically significant results with students, who are roughly the same age, similarly educated, and capable of following instructions, we assume that the protocol will yield better results with participants from a wider population. However, one side-effect of this participant pool is that many participants come from a location other than the culture we were trying to profile. This raises a question on how to correctly screen for a participant from a specific driving culture, i.e., When someone has driven long enough in a specific location to qualify for the experiment.

Over time, when it becomes clearer exactly which measures best capture the differences in behavior and interaction between cultures, the open-ended qualitative observations of researchers can give way to pre-programmed sensors or measures of key variables. These may someday be captured as standard metrics, such as those defined in SAE J2944 [163], and be computed from the data generated from this simulator or instrumented vehicles. Until that time, however, our driving simulator analysis environment needs to allow researchers to play and replay the interactions between the participants and code behaviors or data points they think are notable.

While this system is the first to compare driver-driver interactions across cultures, it is not our intent in this paper to make broad claims about driving cultures.

For future studies using *StrangeLand*, where the goal will be to characterize driving interaction rather than prove the system functions, we will make greater efforts to recruit local participants for the study and be conscientious about how we sample the population. Certainly, there is a wide range of individual variations in driving behavior within a culture. Therefore, we are looking for ways to profile demographic differences within a population to understand how some of these differences interact with the broader norms in each region. In future studies, we would like to perform stimulus sampling [210] by incorporating study runs from three cities in each culture we are trying to profile.

As this is a simulator study, one essential question is how the motions and gestures captured during the experiment align with those that occur during actual driving. Simulation studies have been a mainstay of transportation research for many years and generally yield results that correlate to on-road behavior [127, 140, 182]. However, as Mullen et al. point out, while simulator driving behavior approximates on-road behavior, it does not replicate on-road behavior [155]. Hence, some efforts to make common instrumentation and measures to study on-road driving in situ are also needed to complement this work.

While this system was designed and evaluated before the pandemic, this system could be adapted to be operated remotely with social distancing. For example, participants can be in different rooms, researchers can maintain a six-foot distance from the participants, and both participants and researchers would be required to wear masks.

4.5 Future Work

4.5.1 2022-2023 Implications

Since both the used software and hardware changed significantly since the original implementation. A rewritten version was implemented. This new version follows the same guiding principles we laid out for simulation platforms in the introduction and can let interactions unfold openly. These changes have been discussed in Section 4.2.3. The implications of these changes are summarized here.

The re-implementation also offered us to future-proof the simulator by enabling some desired features. Here is a short list of the most essential features now available:

Any number of participants The simulator is now designed to handle any number of participants, with the current implementation handling up to 6 participants simultaneously. Of course, hardware limitations, especially for the Oculus Quest, apply.

Multiple Spawning Types The original version only was capable of spawning a participant as a driver of a vehicle. The current implementation makes it easy to spawn participants in multiple different types. This could be pedestrians and cyclists that then get to generate whatever virtual representation they need to take part in the simulated environment.

Centralized recording The visualization and analysis of the vehicle data for the previous simulator versions was particularly tricky as the data was distributed over the different participating computers. However, due to the changes in topology and the requirement to support low-performance devices such as Stand-alone VR headsets, the server is the only place that computes the physics simulation and takes direct steering input. This creates one ground truth for the simulated environment that is logged and recorded for later analysis, including ReRun (see next chapter).

The new implementation also requires less hardware (as discussed above) packing and transportation is thus also easier and safer. (See Figure 4.3.)

4.5.2 Pedestrian Study (JiHyun Jeong)

Other researchers have extended the updated version to facilitate their research needs. Cornell Information Science Ph.D. candidate JiHyun Jeong [103] builds on top of *StrangeLand*, adding the ability to spawn participants as AV-passenger or a pedestrian beyond the original option to spawn them as the driver of a vehicle [104, 178, 179]. These new spawn types make use of the new flexible implementation of *StrangeLand* and allow for integration with other systems like the **ZED 2i** from StereoLabs for skeleton tracking.

Figure 4.12: Implementation of an Autonomous Vehicle passenger and Pedestrian inside the *StrangeLand* simulator[104]. *Credit: JiHyun Jeong*

4.5.3 Use as a Research Platform

With this contribution of *StrangeLand*, we demonstrated how multi-user simulation can be used to explore and analyze the interaction among participants in a shared environment. While this example focused on interactions between drivers, the methodological implications and the system itself can also be applied in other contexts and research areas, such as HRI and HCI.

The networked multiplayer approach to handling multiple users makes *StrangeLand* a platform that could be extended to include multiple participants and robots, all joining as network clients into the same shared virtual environment. Interactions between these real and simulated agents could then be recorded in constructed tasks or more open-ended scenarios to discover how behavioral patterns and interactions unfold over time.

The use of such controlled but open-ended scenarios could be highly beneficial for work on HRI and HCI to discover how design choices affect the future deployment of interactive technologies.

4.6 Conclusion

In conclusion, the *StrangeLand* system provides a platform for capturing naturalistic driving interactions between people in different locations. It uses lightweight, portable consumer-grade equipment, and open and extensible software which makes the system suitable for multi-location studies. This also allows other researchers to replicate the methods and run comparison studies with their local population. The system lays the groundwork for future research to profile regional differences in driving culture, essential for drivers and autonomous systems to adapt to.

StrangeLand as a system, especially after its re-implementation(see Section 4.5.1 has been developed to be extensible, to be configured for a variety of scenarios and research questions (see Section 4.5.2). While the early studies deployed the SAGET style methodology longer form studies, remote participation studies, and HRI experiments are all within the capabilities extension of *StrangeLand* can provide.

The pilot studies in this work showed that *StrangeLand* can elicit participant responses consistently and record their reactions both through simulator data streams and questionnaires.

The synchronized recording of all distributed data streams allows for a wide range of analysis methods. The data can be animated for behavioral coding to discover interaction patterns (see Figure 4.6); used for established measures such as steering entropy [101, p.137], or for artificial simulation of driving behaviors.

Continued development on *StrangeLand* aims to incorporate different kinds of research questions, in particular looking at supporting higher numbers of participants and integrating virtual robots, and training of artificial agents.

CHAPTER 5

RERUN

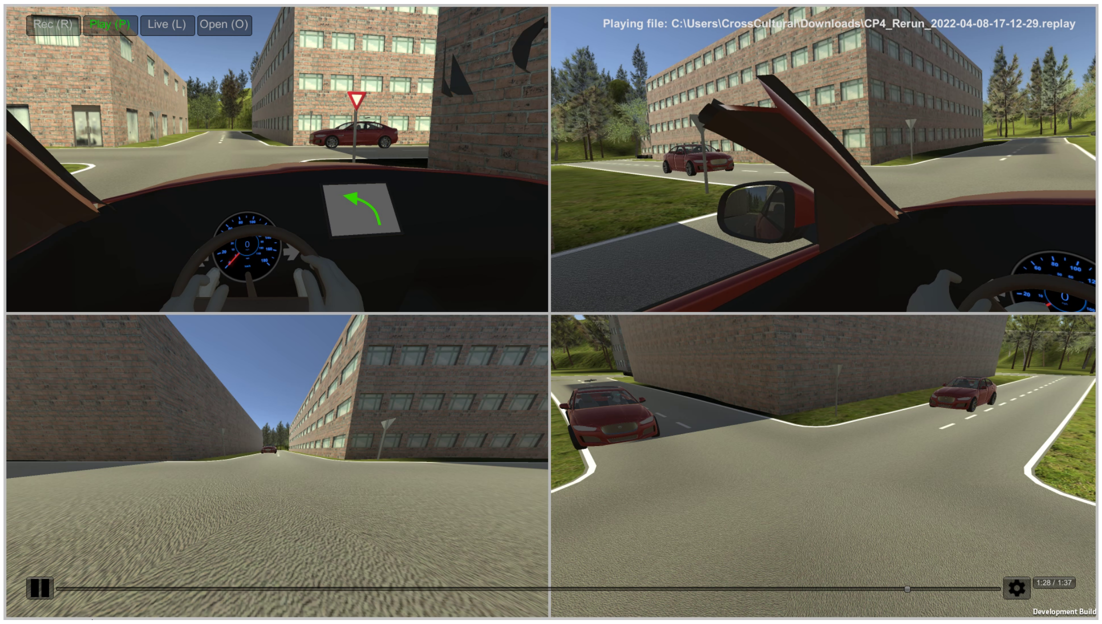


Figure 5.1: Showing four different views inside ReRun of a recorded scenario as it is being played back.

ReRun[1] is a software system to support post-facto analysis in simulation research. While generally application agnostic, the initial version was developed and used for the recording and analysis of the multi-person driving simulator *StrangeLand*, as discussed in the previous section. (See Figure 5.1.)

Just like the simulator itself, ReRun is built in Unity 3D and captures the virtual behavior of participants and their interactions with virtual objects. These recorded behaviors can then be played back from any perspective in the virtual space. This is useful in multi-agent interaction studies, like interactions between drivers or drivers with simulated autonomous vehicles (AVs). Researchers can sift through

[1]This chapter reuses material from the original publications [79]. Those works were co-authored with *Harald Haraldsson, Navit Klein, Lunshi Zhou, Avi Parush, and Wendy Ju*, but I was the lead author of those papers and the primary researcher on that body of work.

scenarios carefully from each participant's perspective or even from an outside observer's perspective.

With this, ReRun revolutionizes the field of qualitative work in VR studies by providing a unique capability that has been largely absent until now. By offering the ability to replay and analyze recorded behaviors from various perspectives, including the first-person ego perspective, ReRun allows researchers to delve into the intricacies of interaction scenarios in a much more personal and immersive manner. This breakthrough feature empowers researchers to gain a deep understanding of the implicit and explicit signaling that occurs between participants and other human or AI-controlled agents within the virtual environment.

5.1 Introduction

With the *StrangeLand* VR simulation platform, we allow researchers to study how drivers will respond to ambiguous scenarios, to novel technologies such as AVs, or to critical scenarios (like cars being broken down on the road) without the risk of staging these situations in the real world [183]. The system is designed to fully simulate all elements separately, meaning that the scenarios can develop and unfold naturally. The complex multi-person simulation presents a great opportunity, as the unfolding interactions can be observed from different points of view, annotated, and behaviorally coded.

5.1.1 System

For this project, we added the functionality of being able to replay simulation runs to the *StrangeLand* simulator [77] (see Chapter 4 and Figure 5.2). Because the *StrangeLand* simulator enables multi-person interaction, the ability to replay driving interactions from different perspectives gives us a greater ability to perform a qualitative analysis of driving interaction and to explore the role that viewpoint plays on perceptions of traffic interaction.

Figure 5.2: Participants using the *StrangeLand* simulator.

ReRun extends the Ultimate Replay 2.0 asset, which is available on the Unity Asset Store[2]. This asset is primarily used by game developers for in-game event replays and offers efficient storage and playback of state-based event data. ReRun is built on top of this asset without modifying its existing code base. This allows for sharing of ReRun code with the research community, which can then be run within projects containing the purchased asset. ReRun implements custom-made features for conveniently recording and playing back VR interactions. The key features from this abstraction layer are ready-to-use prefabs enabling the recording of head pose and hand tracking data, a virtual camera rig for flexible multi-view playbacks, handling of session metadata, and a decoupled desktop UI not visible to the VR participant. The updated version of the *StrangeLand* [77] simulator that

[2]**Ultimate Replay 2.0**: `https://assetstore.unity.com/packages/tools/camera/ultimate-replay-2-0-178602`

integrates tightly with ReRun can be found on GitHub[3]. ReRun is published on GitHub under the MIT license[4].

5.1.2 Related Work

Replaying actions and motions from a virtual environment are a common feature found in many video games, such as Activition's Call of Duty and Codemasters' Grid2. Often, game replays are used to recap important events or even offer players an opportunity to review gameplay from different perspectives.

Recording the entire game state, then, is a feature common to many commercial game engines. Outside of gaming, virtual reality playback has been used to help athletes review their sports performance [18], dentists to gain sensorimotor skill [120], public speakers to understand how their speaking performance might be perceived [219], and VR experience designers to understand the first-person experience that viewers may have of their systems [134].

However, the ability to playback driving simulation experiences is novel. For the most part, driving researchers use event data and virtual telematics data from the simulation environment to determine driving activities and performance. To our knowledge, no previous work has been published demonstrating the ability to playback driving simulation runs for post-facto analysis; this work demonstrates how it enables researchers to revisit interactions from different perspectives so that the effect of, say, distracting events or field of view might have influenced driving interaction outcomes.

[3] **Updated *StrangeLand* simulator:** `https://github.com/FAR-Lab/CrossCulturalDriving2021`

[4] **ReRun:** `https://github.com/FAR-Lab/ReRun`

5.2 Description

To experiment and develop the features, we extend the *StrangeLand* simulator [77] to allow for recording and playback of the vehicle behaviors (car motion, wheel, and steering wheel motion, on-screen driving instructions, indicator lights) as well as the driver's position and hand gestures.

The published video [77] demonstrates how to record and playback these scenarios with different perspectives to form a complete picture of how interactions occur in this simulator.

5.3 Future Work and Implications for Qualitative Research in VR

To discover how interactions unfold, we need appropriate research tools that allow for the reconstruction of all elements that influence these interactions. When it comes to video recordings of interaction unfolding in the real-world many tools, exist, e.g., Elan [126] and Chronoviz [59] that allow for a post-facto analysis. By annotating the video, audio, and other recording streams, researchers can reason through the different signals between participating agents that let to an interaction. Future work on ReRun should incorporate these kinds of annotation features to facilitate this kind of analysis inside scenarios in the simulated virtual space.

VR-recorded interactions also offer additional features for interaction reconstruction. Besides standard features like annotating the time section, the 3d environment can annotate the 3d relationship between the recorded agents. Furthermore, the recording can be played back from any angle, making understanding specific points of view more straightforward.

The hope is that this work can find application in interaction analysis far beyond the use inside of the AV context. Simulating robotic agents for HRI and interfaces for HCI could benefit from both open-ended simulations of scenarios and behavioral analysis through ReRun to discover how interaction takes place.

5.4 Conclusion

In conclusion, *ReRun* extends the *StrangeLand* simulator to allow for recording and playback of vehicle behaviors, driver positions, and hand gestures, providing a comprehensive view of how interactions unfold in the simulator. While demonstrated within *StrangeLand*, *ReRun* more generally is developed as an extension for Unity that can be used to record and playback for multi-agent interaction that is taking place in VR. This tool allows for perspective-taking and behavioral coding to uncover how and why a scenario unfolded in a certain way.

In its initial version, collaborators used it extensively to identify outliers and discover scenarios for further analysis.

Future work on *ReRun* should incorporate annotation features to facilitate more accessible post-facto analysis of interactions in the simulated virtual space. Annotation in 3D space is a yet unexplored space, making the tool's design an open exploration design question. This includes potentially exporting data for further qualitative analysis and the automatic detection of specific interactions.

This work has implications beyond the AV context, as it can be applied to interaction analysis in various fields, including HRI and HCI, by simulating scenarios and analyzing behavioral patterns.

CHAPTER 6

CONCLUSION

As technology keeps developing and we engage more and more with proactive artificial agents, we need new tools that support as designs the process of designing those agents' behaviors.

These research tools will need to help us understand how our design choices will shape the environment. Imagining a possible future through immersive simulation is one way to ensure the design process follows a meaningful direction.

In this thesis, I have explored three different approaches to open-ended simulation that help discover how interaction with autonomous agents can take place. The three different simulators include elements of letting naturalistic uncontrolled elements like behaviors (*VR-OOM*/*StrangeLand*) and environmental factors (*XR-OOM*) play a role in how a scenario places out. And while this work took place within the context of AV research, the implications can be carried over to other research questions concerning HRI and HCI.

6.1 Recapping Orbits of Interaction

The development of different simulators throughout this thesis has provided valuable opportunities to explore various interactions with AVs as an first intelligent agent acting in our world. Each simulator offers distinct advantages and focuses on other aspects of the AV experience. By leveraging these simulators, this thesis aims to advance and enhance our understanding of the complexities and possibilities of interactions within these domains.

As the initial implementation, the ***VR-OOM*** simulator leveraged the immersive feeling of driving in a car while replacing the visual component of a passenger-riding experience. This simulator proved particularly valuable in testing critical and more controlled scenarios. With the introduction of an updated tracking version, we were able to incorporate Digital Twin systems, enabling participants to explore a digital replica of an environment in virtual reality while actively navigating through it. As a result, *VR-OOM* stands out as the most controlled simulator for individual participants within the context of this thesis work, allowing us to delve into the inner circle of interaction and gain insights into the intricacies of user experiences.

XR-OOM, an extension of the *VR-OOM* simulator, goes a step further by letting the real world literally pass through to the participant's vision and integrating the natural traffic environment into the testing process. While still designed for one participant at a time, *XR-OOM* offers a rich experience incorporating the real-world context of traffic scenarios. The visual augmentation provided by the system allows for the construction of specific scenarios while maintaining a visually and sensorially immersive environment. As a result, this simulator is handy for exploring interactions and observing complex designs and behaviors in action.

Lastly, ***StrangeLand*** represents the outermost orbit of interaction research, focusing on understanding interactions with other road users. As a flexible research platform, it can be extended to include different types of road users and incorporates the use of **ReRun** for behavioral analysis. This outermost orbit enables a comprehensive understanding of how interactions occur and provides insights into the cues that participants attend to. With the ability to conduct behavioral coding, researchers can gain a detailed knowledge of AV behaviors and the behaviors of other autonomous systems. The versatility of *StrangeLand* allows for the seamless integration of different methods, such as the Robot Operating System, to simulate and study realistic AV software or interactions with other types of robots.

Collectively, these different simulator methods enable the gathering of data on participant behaviors and interactions with each other and the Autonomous System, paving the way for improved design and development of autonomous technologies.

Finally, all published simulators adhere to the *Research Platform* guidelines in Section 1.2.

Accessibility For each simulator, I published the system details like components and hardware, the study protocols used to run the study, including the used questionnaires, Standard Operating Procedures (e.g., Appendix A.2.3), and finally, publications included example data as videos and excepted to demonstrate the capabilities of the system.

Usability All publications included proof of concept demonstration and studies that showed the system working and generating consistent data.

Adaptability The simulators are built in Unity and heavily make use of prefabs and separate game objects, with various relationships established through Unity's inspector, making the implementation easily modifiable.

Extendable Especially for the simulator *StrangeLand*, we took good care at writing the code in such a way as to make it easily extendable while still providing valuable features and tools. All simulators are published under the MIT license. (See Appendix A.1 for links to the different projects' repositories.)

6.1.1 The Serendipity of Complex Scenarios

The value of open-ended immersive simulation lies in its ability to foster the discovery of novel combinations of scenarios, technology, and human interactions. Unlike traditional, tightly controlled scenarios used in lab studies, open-ended simulation enables a more serendipitous exploration of human-robot interactions. This approach allows researchers to capture a broader range of experiences, uncover unexpected uses of technology, and observe interactions that might not have been anticipated.

While some open methods exist, such as instrumented free real-world driving, many immersive driving simulators still lack the naturalistic unfolding of scenarios. Open-ended immersive simulation bridges this gap, providing a platform where scenarios can organically evolve, reflecting real-world complexity and variability. By embracing this more open approach, researchers gain the opportunity to observe authentic interactions in their natural context, leading to deeper insights and a more comprehensive understanding of human-robot dynamics.

The different simulators presented in this thesis embody these ideas to varying degrees. Below we assess the different simulators' strengths and weaknesses regarding this dimension: (see Figure 6.1)

StrangeLand makes use of controlled scenarios. The starting points and instructions for the participants are predefined and fixed. That said, the way the scenario unfolds, how participants interpret the instructions, and interaction with other drivers lead to unforeseen outcomes.

VR-OOM As this method is deployed on the road, in real-world traffic, it inevitably exposes the participant to unforeseen scenarios. Still, all of the visual

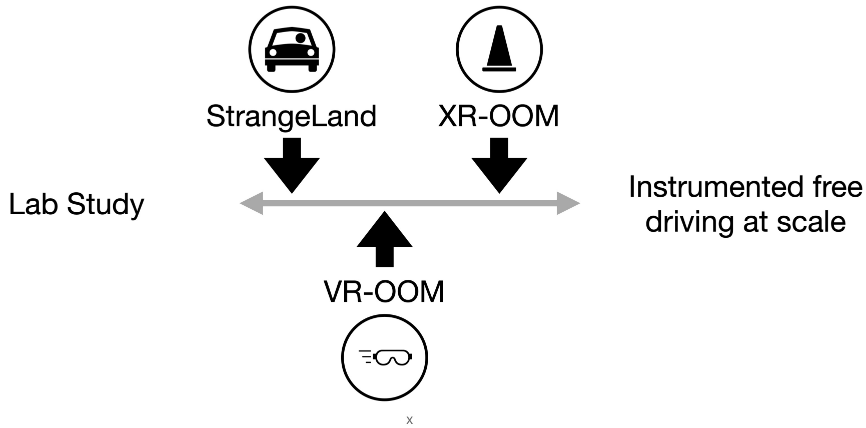


Figure 6.1: On a scale between *Lab studies* (constrained) and *Instrumented free driving at scale* (open). Each simulator has varying degrees of openness they support. This graphic visualizes were approximately the simulators fall on this dimension.

elements seen by the participant are simulated, giving the researcher a significant amount of control over the scenario.

XR-OOM is by the simulator most affected by outside world influence, especially if a participant is not driving the vehicle but is a passenger as a research driver through regular traffic.

6.1.2 Technological Challenges and Opportunities

VR technology is only in its 3rd generation [185]. The considerable improvements that have taken place since the development of the first consumer headset (Oculus CV1) to the recently released Meta Quest Pro point towards a field that will continue to see more improvements. Even if the graphics and tracking quality only continue to experience marginal improvements, VR-based simulators will continue to become more immersive and easier to deploy.

In regards to rendering features, it would be great to see improvements in augmented reality glasses that include more contextual awareness for both light and sound. Features that have now been shown to be available in their initial version of Apples Vision Pro headset [10]. One particular use case for *XR-OOM* would be the correct casting of sunlight and skylight on virtual objects and calculating and superimposing correct shadows cast by these virtual objects. These kinds of technical advancements could be easily prototypes and explored as the next research step to technically improve the *XR-OOM* system. Other aspects, such as model fidelity and shader quality, are already very sophisticated.

Besides these rendering techniques, there are many other additional technical challenges to situating virtual elements seamlessly into the view.

Occlusion is essential to place virtual objects into the real-world view properly. Especially real-world objects with a lot of details, like trees, present a particular challenge here. As new computer vision capabilities are being developed, combinations of methods (e.g., spatial mapping and image segmentation [115]) might be able to handle this kind of challenge.

Object tracking of real-world elements like cars on the road would provide new capabilities for constructing scenarios. An existing traffic scenario could be detected and augmented or used for an experiment.

Augmentation of parts of objects. Currently, many entire objects like geese, pedestrians, and cars are augmented into the view. However, that ability to only replace or modify parts of an object would give in particular *XR-OOM* the ability to expand the types of studies quickly. E.g., another vehicle in traffic could be made to look like an autonomous vehicle.

6.1.3 Building Models of Interaction

The simulators developed in this thesis aim to generate a rich stream of data that can be leveraged to build models of human behavior. Understanding how people behave and interact with systems is crucial in designing effective behaviors for autonomous systems. Moving forward, it is essential to systematically expand the range of data collected and integrate additional sensor systems, such as cameras, to track reactions and interactions more comprehensively. In particular, the *XR-OOM* simulator (see Figure 6.2) aims to incorporate additional systems to construct complete traffic scenarios and analyze the specific details that influence system behavior. Ultimately, modeling behavior and studying the interaction with autonomous vehicles are pivotal in informing the design of future behaviors for these systems.

Figure 6.2: Capturing the participant with a regular camera allows us to measure the participant's motion and position.

6.1.4 Interaction with Automation

The simulations presented in this thesis are tools to explore how the interaction between people and AVs takes place and unfolds. Even though these simulators are tested in this specific regime of automation research, the findings and some of the systems can be extended to answer other (behavior) design questions for systems that are being automated. In the following sections, these possibilities are briefly sketched out to explore how some HRI and HCI interaction design challenges could be tackled using the developed simulator concepts.

Human Robot Interaction

The field of HRI has addressed many different challenges in designing robots, their behaviors, and their appearance. The ACM/IEEE 'International Conference on Human-Robot Interaction' was established in 2006 and explores these kinds of research questions. Within this community, using immersive simulation is not a new approach [51, 132, 193, 212]. However, specific open-ended simulations such as the one presented in this thesis could be helpful for research questions within the HRI context. I explored a few options on how this approach could be used for HRI research questions.

Implicit Interaction. In social spaces, robots will need to use their existing modalities like motion [184] or sound [152, 153] to facilitate smooth and implicit interactions with the people in that space. The integration of robotic control systems like ROS [170] in immersive simulations would give *behavior designs* the opportunity to quickly iterate through different ideas. The direct integration with control code would mean that the finished designs can be easily transferred to a real-world robot.

Human Robot Teaming. In shared social spaces [195], understanding how teamwork [54, 149] unfolds as people and robots collaborate is key in optimizing the teaming strategy. Reconstructing these interactions requires researchers to record and playback the event streams to allow for qualitative analysis. To discover how cues are shared and used between agents (robots and people) to facilitate interaction, simulations, in particular **ReRun**, offers us unique capabilities to record *all* dynamic elements of a virtual environment. Making it a perfect tool to uncover how interactions unfold.

Learning from Demonstration. One prominent approach in HRI is to design robot behavior using demonstrations by a human (e.g. [11, 39]). The use of simulated robots in a virtual reality space for this approach has multiple benefits. First, it allows the participant to be immersed in the environment in which they are demonstrating the motion, as compared to making these recordings in a lab setting. Second, using VR for the recording of such demonstrations makes the data collection process easier. Instead of having to transport an entire robot system, a VR headset (e.g., Meta Quest 2) with the appropriate software can be used (compare Section 4.2.3)

Wizard of Oz (WoZ) methods have been in use within HRI for a long time (e.g. [214]). Combining WoZ methods with a shared virtual reality space like ***StrangeLand*** would allow for new possibilities for scenario construction and wizarding. This is because the researcher could fully embody and appear as the robot agent in the shared virtual space to act out scenarios. This can be easily achieved by mapping their tracked body to the motions of the virtual robot.

Human Computer Interaction

HCI similarly has challenges that could be addressed within immersive VR applications.

Simulating Context. Within End-User applications reacting to context is an essential design challenge (e.g. [169]). The construction of different virtual reality contexts is an inherent feature of this technology. The combination of prototyping methods for End-User applications with VR allows for flexible context-aware prototyping of interfaces.

Industrial Machine Operation. Immersive VR simulation could be used for both interface design tasks as well as for the training of operators. With modern XR headsets, this application could also include real-world interfaces while simulating the dangerous elements of a system in its context.

These are just a few examples of HCI challenges in which the explored methods in this thesis could find application.

6.2 Reception of the Work

Since the original publication of *VR-OOM* in 2018, *XR-OOM*, and *StrangeLand* in 2022, the work presented in this thesis has gained tracking within the academic research community.

6.2.1 Research Pickup at Cornell

For the different projects, different follow projects were picked up and have been addressed in the respective sections.

VR-OOM was picked up by Yavo-Ayalon et al. [215] as described in Section 2.2.1 in which *VR-OOM* was to investigate if "Situated and shared experiences can motivate community members to plan shared action, promoting community engagement."

XR-OOM was integrated with the *Portabello* as shown in Section 3.7.4 by Fanjun et al. [53] to improve tracking accuracy and speed with a mapped reference.

StrangeLand got extended to include other traffic participants in a three-way traffic interaction scenario see Section 4.5.2.

6.2.2 Research Pick-up Internationally

Through its adoption, my work has contributed to work in AV research and Human-Robot-Interaction. Here are some of the most important works that cite the work from my thesis:

VR-OOM Sportillo, Daniele, Alexis Paljic, and Luciano Ojeda. **"Get ready for automated driving using virtual reality."** Accident Analysis & Prevention 118 (2018): 102-113.

VR-OOM Gao, Yuxiang, and Chien-Ming Huang. **"Evaluation of socially-aware robot navigation."** Frontiers in Robotics and AI 8 (2022): 420.

VR-OOM Detjen, Henrik, Bastian Pfleging, and Stefan Schneegass. **"A wizard of oz field study to understand non-driving-related activities, trust, and acceptance of automated vehicles."** In 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 19-29. 2020.

XR-OOM Wilson, Graham, Mark McGill, Daniel Medeiros, and Stephen Brewster. **"A Lack of Restraint: Comparing Virtual Reality Interaction Techniques for Constrained Transport Seating."** IEEE Transactions on Visualization and Computer Graphics 29, no. 5 (2023): 2390-2400.

XR-OOM McGill, Mark, Graham Wilson, Daniel Medeiros, and Stephen Anthony Brewster. **"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, pp. 1-15. 2022.

StrangeLand Lee, Seong Hee, Nicholas Britten, Avram Block, Aryaman Pandya, Malte F. Jung, and Paul Schmitt. **"Coming In! Communicating Lane Change Intent in Autonomous Vehicles."** In Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, pp. 394-398. 2023.Gao, Yuxiang, and Chien-Ming Huang. **"Evaluation of socially-aware robot navigation."**

StrangeLand Nai, Wei, Zan Yang, Yinzhen Wei, Jierui Sang, Jialu Wang, Zhou Wang, and Peiyu Mo. **"A comprehensive review of driving style evaluation approaches and product designs applied to vehicle usage-based insurance."** Sustainability 14, no. 13 (2022): 7705.

6.3 Closing

In conclusion, this thesis has highlighted the significance of driving simulation environments in the field of automotive design and human factors engineering. The three simulators offer a controlled and replicable setting to assess the impacts of interfaces, interactions, and intelligent behaviors exhibited by self-directed systems. By exploring new and immersive simulation technologies, we have created the ability to explore the dynamics of trust, teamwork, and interaction within the context of Autonomous Vehicles.

This research has examined various orbits of AV interaction, spanning from the interior of the vehicle to the broader traffic environment, enabling researchers and car manufacturers to understand how trust, resolution of ambiguous situations, and the evolving role of human control change with their designs. The insights gained from this work contribute to a much deeper understanding of how simulation can enhance our understanding of HRI and HCI interactions and inform the design of future autonomous systems. This has already been seen by the pickup from the research community.

BIBLIOGRAPHY

[1] perceived_barriers_to_stem_higher_education_aera_version.pdf.

[2] Kareem Abdelgawad, Jürgen Gausemeier, Roman Dumitrescu, Michael Grafe, Jörg Stöcklein, and Jan Berssenbrügge. Networked driving simulation: Applications, state of the art, and design considerations. *Designs*, 1(1): 4, 2017.

[3] Talal Al-Shihabi and Ronald Mourant. Toward more realistic driving behavior models for autonomous vehicles in driving simulators. *Transportation Research Record: Journal of the Transportation Research Board*, 1843(1843):41–49, 2003.

[4] R Wade Allen, Theodore J Rosenthal, and Marcia L Cook. *A short history of driving simulation*, chapter 2, pages 2–1 to 2–16. CRC press, Boca Raton, 2011.

[5] Micah Alpern and Katie Minardo. Developing a car gesture interface for use as a secondary task. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 932–933. ACM, 2003.

[6] Ignacio Alvarez, Laura Rumbel, and Robert Adams. Skyline: a rapid prototyping driving simulator for user experience. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 101–108. ACM, 2015.

[7] Ignacio Alvarez, Jennifer Healey, and Erica Lewis. The skynivi experience: evoking startle and frustration in dyads and single drivers. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 76–81. IEEE, 2019.

[8] Sonia Amado, Elvan Arıkan, Gülin Kaça, Mehmet Koyuncu, and B Nilay

Turkan. How accurately do drivers evaluate their own driving behavior? an on-road observational study. *Accident Analysis & Prevention*, 63:65–73, 2014.

[9] Kaarin J. Anstey, Joanne Wood, Stephen Lord, and Janine G. Walker. Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review*, 25(1):45 – 65, 2005. ISSN 0272-7358. doi: https://doi.org/10.1016/j.cpr.2004.07.008. URL `http://www.sciencedirect.com/science/article/pii/S0272735804001345`.

[10] Apple. Introducing vision pro, may 2023. URL `https://www.apple.com/apple-vision-pro/`.

[11] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

[12] Bruce Artz, Larry Cathey, Peter Grant, Dan Houston, Jeff Greenberg, and Max Mariani. The design and construction of the visual subsystem for virttex, the driving simulator at the ford research laboratories. In *Driving simulation conference*, pages 255–262, 2001.

[13] Ed Bachelder. Helicopter aircrew training using fused reality. Technical report, Systems Technology Inc., 2006.

[14] Sonia Baltodano, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju. The rrads platform: A real road autonomous driving simulator. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '15, page 281–288, New York, NY, USA, 2015. Association for Computing Machinery. ISBN

9781450337366. doi: 10.1145/2799250.2799288. URL https://doi.org/10.1145/2799250.2799288.

[15] Sonia Baltodano, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju. RRADS: real road autonomous driving simulation. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 283–283. ACM, 2015. URL http://dl.acm.org/citation.cfm?id=2702099.

[16] Sonia Baltodano, Jesus Garcia-Mancilla, and Wendy Ju. Eliciting driver stress using naturalistic driving scenarios on real roads. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 298–309. ACM, 2018.

[17] Salvador Bayarri, Marcos Fernandez, and Mariano Perez. Virtual reality for driving simulation. *Communications of the ACM*, 39(5):72–76, 1996.

[18] Benoit Bideau, Richard Kulpa, Nicolas Vignais, Sébastien Brault, Franck Multon, and Cathy Craig. Using virtual reality to analyze sports performance. *IEEE Computer Graphics and Applications*, 30(2):14–21, 2009.

[19] Miles Branman. *Honda's Dream Drive VR experience may be the future of in-car entertainment*, 2017 (accessed September 17, 2017). URL https://www.digitaltrends.com/cars/ces-2017-honda-dream-drive/.

[20] Alexandra W.D. Bremers, Ali Özgür Yöntem, Kun Li, Daping Chu, Valerian Meijering, and Christian P. Janssen. Perception of perspective in augmented reality head-up displays. *International Journal of Human-Computer Studies*, 155:102693, 2021. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2021.

102693. URL `https://www.sciencedirect.com/science/article/pii/S1071581921001117`.

[21] Heiner Bubb, Klaus Bengler, Rainer E. Grünen, and Mark Vollrath. *Automobilergonomie*. ATZ/MTZ-Fachbuch. Springer Fachmedien Wiesbaden, Wiesbaden, 2015.

[22] GE Burnett, A Irune, and A Mowforth. Driving simulator sickness and validity: How important is it to use real car cabins. *Advances in transportation studies*, 13:33–42, 2007.

[23] Alessandro Oronzo Caffò, Luigi Tinella, Antonella Lopez, Giuseppina Spano, Ylenia Massaro, Andrea Lisi, Fabrizio Stasolla, Roberto Catanesi, Francesco Nardulli, Ignazio Grattagliano, et al. The drives for driving simulation: A scientometric analysis and a selective review of reviews on simulated driving research. *Frontiers in Psychology*, 11, 2020.

[24] Fanta Camara, Nicola Bellotto, Serhan Cosar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Gustav Markkula, et al. Pedestrian models for autonomous driving part ii: high-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[25] Shi Cao, Krithika Nandakumar, Raiju Babu, and Benjamin Thompson. Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display. *Virtual Reality*, 24(3):503–513, 2020.

[26] Philip Cash. Where next for design research? understanding research impact and theory building. *Design Studies*, 68:113–141, 2020. ISSN 0142-694X.

doi: https://doi.org/10.1016/j.destud.2020.03.001. URL https://www.sciencedirect.com/science/article/pii/S0142694X20300211.

[27] Matthew Chinman, Joie Acosta, Patricia Ebener, and Amy Shearer. "what we have here, is a failure to [replicate]": Ways to solve a replication crisis in implementation science. *Prevention Science*, 23(5):739–750, 2022. doi: 10.1007/s11121-021-01286-9. URL https://doi.org/10.1007/s11121-021-01286-9.

[28] Mark Claypool and Kajal Claypool. Latency and player actions in online games. *Commun. ACM*, 49(11):40–45, November 2006. ISSN 0001-0782. doi: 10.1145/1167838.1167860. URL https://doi.org/10.1145/1167838.1167860.

[29] Mark Colley, Pascal Jansen, Enrico Rukzio, and Jan Gugenheimer. 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), dec 2022. doi: 10.1145/3494968. URL https://doi.org/10.1145/3494968.

[30] Open Simcraft Community. Open source motion simulation, July 2023. URL https://opensimcraft.github.io/.

[31] State Farm Mutual Automobile Insurance Company. Teen driving 101: a step-by-step test of essential driving skills, unknown date. URL https://www.statefarm.com/simple-insights/auto-and-vehicles/teen-driving-101-a-stepbystep-test-of-essential-skills.

[32] Monroe Conner. *Fused Reality: Making the Imagined Seem Real*, 2015 (ac-

cessed September 17, 2017). URL https://www.nasa.gov/centers/armstrong/features/fused_reality.html.

[33] BMW Corporation. The bmw driving simulation centre: Simulating the future of mobility today, July 2022. URL https://www.bmw.com/en/events/nextgen/driving-simulator.html.

[34] Lee J Cronbach and Paul E Meehl. Construct validity in psychological tests. *Psychological bulletin*, 52(4):281, 1955.

[35] Eduardo Cuervo, Krishna Chintalapudi, and Manikanta Kotaru. Creating the perfect illusion: What will it take to create life-like virtual reality headsets? In *Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications*, HotMobile '18, page 7–12, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356305. doi: 10.1145/3177102.3177115. URL https://doi.org/10.1145/3177102.3177115.

[36] W Dale Dannefer. Driving and symbolic interaction 1. *Sociological Inquiry*, 47 (1):33–38, 1977.

[37] Santiago Higuera de Frutos and María Castro. Assessing sim racing software for low-cost driving simulator to road geometric research. *Transportation Research Procedia*, 58:575–582, 2021. ISSN 2352-1465. doi: https://doi.org/10.1016/j.trpro.2021.11.076. URL https://www.sciencedirect.com/science/article/pii/S2352146521008358. XIV Conference on Transport Engineering, CIT2021.

[38] Joost De Winter, Peter M van Leeuwen, and Riender Happee. Advantages

and disadvantages of driving simulators: A discussion. In *Proceedings of measuring behavior*, volume 2012, page 8th. Citeseer, 2012.

[39] Jan de Wit, Paul Vogt, and Emiel Krahmer. The design and observed effects of robot-performed manual gestures: A systematic review. *J. Hum.-Robot Interact.*, 12(1), feb 2023. doi: 10.1145/3549530. URL https://doi.org/10.1145/3549530.

[40] Jerry L Deffenbacher. Anger, aggression, and risky behavior on the road: A preliminary study of urban and rural differences 1. *Journal of Applied Social Psychology*, 38(1):22–36, 2008.

[41] Jerry L Deffenbacher, Rebekah S Lynch, Eugene R Oetting, and Randall C Swaim. The driving anger expression inventory: A measure of how people express their anger on the road. *Behaviour research and therapy*, 40(6):717–737, 2002.

[42] Forward Development. City car driving [pc game]. version 1.5, 2017. URL http://www.citycardriving.com/.

[43] Debargha Dey and Jacques Terken. Pedestrian interaction with vehicles: roles of explicit and implicit communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 109–113. ACM, 2017.

[44] Piotr Didyk, Elmar Eisemann, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Perceptually-motivated real-time temporal upsampling of 3d content for high-refresh-rate displays. In *Computer Graphics Forum*, volume 29, pages 713–722. Wiley Online Library, 2010.

[45] Tilman Dingler, Kai Kunze, and Benjamin Outram. Vr reading uis: Assessing text parameters for reading in vr. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–6, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356213. doi: 10.1145/3170427.3188695. URL `https://doi.org/10.1145/3170427.3188695`.

[46] Thomas A Dingus, Sheila G Klauer, Vicki L Neale, A Petersen, Suzanne E Lee, JD Sudweeks, MA Perez, J Hankey, DJ Ramsey, S Gupta, et al. The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment. Technical report, U.S. Department of Transportation, National Highway Traffic Safety Administration, 2006.

[47] Joshua E Domeyer, John D Lee, Heishiro Toyoda, Bruce Mehler, and Bryan Reimer. Interdependence in vehicle-pedestrian encounters and its implications for vehicle automation. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[48] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.

[49] Florian Echtler and Maximilian Häußler. Open source, open science, and the replication crisis in hci. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, page 1–8, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356213. doi: 10.1145/3170427.3188395. URL `https://doi.org/10.1145/3170427.3188395`.

[50] Mica R Endsley. Situation awareness global assessment technique (sagat). In

Proceedings of the IEEE 1988 national aerospace and electronics conference, pages 789–795. IEEE, 1988.

[51] Roberta Etzi, Siyuan Huang, Giulia Wally Scurati, Shilei Lyu, Francesco Ferrise, Alberto Gallace, Andrea Gaggioli, Alice Chirico, Marina Carulli, and Monica Bordegoni. Using virtual reality to test human-robot interaction during a collaborative task. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 59179, page V001T02A080. American Society of Mechanical Engineers, 2019.

[52] Roni Factor, David Mahalel, and Gad Yair. The social accident: A theoretical model and a research agenda for studying the influence of social and cultural characteristics on motor vehicle accidents. *Accident Analysis & Prevention*, 39 (5):914–921, 2007.

[53] Bu Fanjun, Stacey Li, David Goedicke, Mark Colley, Gyanendra Sharma, and Wendy Ju. Portobello: Extending driving simulation from the lab to the road. *Manuscript submitted for publication.*, 2023.

[54] Pooyan Fazli and Alan K. Mackworth. Human-robot teams in entertainment and other everyday scenarios. *CoRR*, abs/0908.2661, 2009. URL `http://arxiv.org/abs/0908.2661`.

[55] Sebastian S. Feger, Sünje Dallmeier-Tiessen, Paweł W. Woźniak, and Albrecht Schmidt. The role of hci in reproducible science: Understanding, supporting and motivating core practices. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359719. doi: 10.1145/3290607.3312905. URL `https://doi.org/10.1145/3290607.3312905`.

[56] Alexander Feierle, Michael Rettenmaier, Florian Zeitlmeir, and Klaus Bengler. Multi-vehicle simulation in urban automated driving: technical implementation and added benefit. *Information*, 11(5):272, 2020.

[57] Donald L Fisher, Jeff K Caird, Matthew Rizzo, and John D Lee. Handbook of driving simulation for engineering, medicine and psychology: an overview. *Handbook of driving simulation for engineering, medicine, and psychology*, pages 1–1 to 1–14, 2011.

[58] Frank Flemisch, Matthias Heesen, Tobias Hesse, Johann Kelsch, Anna Schieben, and Johannes Beller. Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work*, 14(1):3–18, 2012.

[59] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. Chronoviz: A system for supporting navigation of time-coded data. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, page 299–304, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450302685. doi: 10.1145/1979742.1979706. URL `https://doi.org/10.1145/1979742.1979706`.

[60] J. S. Freeman, G. Watson, Y. E. Papelis, T. C. Lin, A. Tayyab, R. A. Romano, and J. G. Kuhl. The iowa driving simulator: An implementation and application overview. In *International Congress & Exposition*. SAE International, feb 1995. doi: https://doi.org/10.4271/950174. URL `https://doi.org/10.4271/950174`.

[61] Lex Fridman, Daniel E Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Aleksandr Patsekin, Julia Kindels-

berger, Li Ding, et al. Mit advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation. *IEEE Access*, 7:102021–102038, 2019.

[62] Max Friedrich, Daniel Nause, Matthias Heesen, Alexandra Keich, Johann Kelsch, Martin Baumann, and Mark Vollrath. Validation of the mosaic-driving simulator–investigating the impact of a human driver on cooperative driving behavior in an experimental simulation setup. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 2052–2056. SAGE Publications Sage CA: Los Angeles, CA, 2013.

[63] Philippe Fuchs. *Virtual reality headsets-a theoretical and pragmatic approach*. CRC Press, 2017.

[64] Kugamoorthy Gajananan, Alfredo Nantes, Marc Miska, Arturo Nakasone, and Helmut Prendinger. An experimental space for conducting controlled driving behavior studies based on a multiuser networked 3d virtual environment and the scenario markup language. *IEEE Transactions on Human-Machine Systems*, 43(4):345–358, 2013. doi: 10.1109/TSMC.2013.2265876.

[65] Epic Games. Unreal game engine 4, mar 2014. URL `http://unrealengine.com`.

[66] Yan Ge, Weina Qu, Caihong Jiang, Feng Du, Xianghong Sun, and Kan Zhang. The effect of stress and personality on dangerous driving behavior among chinese drivers. *Accident Analysis & Prevention*, 73:34–40, 2014.

[67] G Geiser. Man-machine communication in motor vehicles, Feb 1985.

[68] GENIVI Alliance. Genivi driving simulator, 2016. URL `https://github.com/GENIVI/genivi-vehicle-simulator`.

[69] Michael A Gerber, Ronald Schroeter, and Julia Vehns. A video-based automated driving simulator for automotive ui prototyping, ux and behaviour research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 14–23, 2019.

[70] Petra Geutner, Frank Steffens, and Dietrich Manstetten. Design of the vico spoken dialogue system: Evaluation of user expectations by wizard-of-oz experiments. In *LREC*, 2002.

[71] Florin-Timotei Ghiurãu, Mehmet Aydın Baytaş, and Casper Wickman. Arcar: On-road driving in mixed reality by volvo cars. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20 Adjunct, page 62–64, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375153. doi: 10.1145/3379350.3416186. URL `https://doi.org/10.1145/3379350.3416186`.

[72] Jonathan M. Gitlin. There's an $80 million driving simulator in iowa, and we tried it out, Mar 2020. URL `https://arstechnica.com/cars/2020/03/theres-an-80-million-driving-simulator-in-iowa-and-we-tried-it-out/`.

[73] D. Goedicke. On-road virtual reality driving simulator, December 2017. URL `http://essay.utwente.nl/74239/`.

[74] David Goedicke, Jamy Li, Vanessa Evers, and Wendy Ju. VR-OOM: Virtual reality on-road driving simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 165. ACM, 2018.

[75] David Goedicke, Alexandra W.D. Bremers, Hiroshi Yasuda, and Wendy Ju. Xr-oom: Mixing virtual driving simulation with real cars and environ-

ments safely. In *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '21 Adjunct, page 67–70, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450386418. doi: 10.1145/3473682.3480266. URL `https://doi.org/10.1145/3473682.3480266`.

[76] David Goedicke, Alexandra W.D. Bremers, Sam Lee, Fanjun Bu, Hiroshi Yasuda, and Wendy Ju. 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*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3517704. URL `https://doi.org/10.1145/3491102.3517704`.

[77] David Goedicke, Carmel Zolkov, Natalie Friedman, Talia Wise, Avi Parush, and Wendy Ju. Strangers in a strange land: New experimental system for understanding driving culture using vr. *IEEE Transactions on Vehicular Technology*, 71(4):3399–3413, 2022. doi: 10.1109/TVT.2022.3152611.

[78] David Goedicke, Mark Colley, Sebastian S. Feger, Michael Goedicke, Bastian Pfleging, and Wendy Ju. Towards sustainable research data management in human-computer interaction, 2023.

[79] David Goedicke, Harald Haraldsson, Navit Klein, Lunshi Zhou, Avi Parush, and Wendy Ju. Rerun: Enabling multi-perspective analysis of driving interaction in vr. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, page 889–890, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399708. doi: 10.1145/3568294.3580211. URL `https://doi.org/10.1145/3568294.3580211`.

[80] Christian Gold, Lutz Lorenz, Daniel Damböck, and Klaus Bengler. Partially automated driving as a fallback level of high automation. *Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren*, 28(29.11):2013, 2013.

[81] Gov.UK. The highway code, road safety and vehicle rules, 2009. URL `https://www.gov.uk/browse/driving/highway-code-road-safety`.

[82] Peter Grant, Bruce Artz, Jeff Greenberg, and Larry Cathey. Motion characteristics of the virttex motion system. In *Proceedings of the 1st human-centered transportation simulation conference*, pages 4–7, 2001.

[83] D Grothkopp, W Krautter, B Grothkopp, F Steffens, and F Geutner. Using a driving simulator to perform a wizard-of-oz experiment on speech-controlled driver information systems. In *Proceedings of the 1st Human-Centered Transportation Simulation Conference*, 2001.

[84] Cameron Grout, William Rogers, Mark Apperley, and Steve Jones. Reading text in an immersive head-mounted display: An investigation into displaying desktop interfaces in a 3d virtual environment. In *Proceedings of the 15th New Zealand Conference on Human-Computer Interaction*, CHINZ 2015, page 9–16, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336703. doi: 10.1145/2808047.2808055. URL `https://doi.org/10.1145/2808047.2808055`.

[85] David Hallvig, Anna Anund, Carina Fors, Göran Kecklund, Johan G Karlsson, Mattias Wahde, and Torbjörn kerstedt. Sleepy driving on the real road and in the simulator–a comparison. *Accident Analysis & Prevention*, 50:44–50, 2013.

[86] PA Hancock and SN De Ridder. Behavioural accident avoidance science: understanding response in collision incipient conditions. *Ergonomics*, 46(12): 1111–1135, 2003.

[87] Peter A. Hancock, Mary Lesch, and Lisa Simmons. The distraction effects of phone use during a crucial driving maneuver. *Accident Analysis & Prevention*, 35(4):501–514, 2003.

[88] Xavier Harding. Pitt stop: Inside Uber's driverless car experiment. *Popular Science*, Sep 2016. URL `https://www.popsci.com/uber-raffi-krikorian-driverless-car/`.

[89] Rebecca L Hartman and Marilyn A Huestis. Cannabis effects on driving skills. *Clinical chemistry*, 59(3):478–492, 2013.

[90] Mathias Heesen, Martin Baumann, Johann Kelsch, Daniel Nause, and Max Friedrich. Investigation of cooperative driving behaviour during lane change in a multi-driver simulation environment. In *Human Factors and Ergonomics Society (HFES) Europe Chapter Conference*, pages 305–318, Toulouse, France, 2012. Human Factors and Ergonomics Society (HFES).

[91] Arne Helland, Gunnar D Jenssen, Lone-Eirin Lervg, Andreas Austgulen Westin, Terje Moen, Kristian Sakshaug, Stian Lydersen, Jørg Mørland, and Lars Slørdal. Comparison of driving simulator performance with real driving after alcohol intake: A randomised, single blind, placebo-controlled, cross-over trial. *Accident Analysis & Prevention*, 53:9–16, 2013.

[92] Wolfgang Hess, Damon Kohler, Holger Rapp, and Daniel Andor. Real-time loop closure in 2d lidar slam. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1271–1278, 2016.

[93] Philipp Hock, Sebastian Benedikter, Jan Gugenheimer, and Enrico Rukzio. CarVR. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 4034–4044, 2017. ISBN 9781450346559. doi: 10.1145/3025453.3025665. URL `http://dl.acm.org/citation.cfm?doid=3025453.3025665`.

[94] Philipp Hock, Johannes Kraus, Franziska Babel, Marcel Walch, Enrico Rukzio, and Martin Baumann. How to design valid simulator studies for investigating user experience in automated driving: review and hands-on considerations. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 105–117. Association for Computing Machinery, 2018.

[95] Philipp Hock, Mark Colley, Ali Askari, Tobias Wagner, Martin Baumann, and Enrico Rukzio. 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*, pages 204–214, 2022.

[96] M d Houtenbos, Joost CF de Winter, AR Hale, PA Wieringa, and MP Hagenzieker. Concurrent audio-visual feedback for supporting drivers at intersections: A study using two linked driving simulators. *Applied ergonomics*, 60: 30–42, 2017.

[97] F. Hu, Y. Deng, W. Saad, M. Bennis, and A. H. Aghvami. Cellular-connected wireless virtual reality: Requirements, challenges, and solutions. *IEEE Communications Magazine*, 58(5):105–111, 2020.

[98] Karl Iagnemma. Supervisory control of vehicles, May 28 2019. US Patent 10,303,166.

[99] Quinate Chioma Ihemedu-Steinke, Demet Sirim, Rainer Erbach, Prashanth Halady, and Gerrit Meixner. Development and evaluation of a virtual reality driving simulator. *Mensch und Computer 2015–Workshopband*, 2015.

[100] Quinate Chioma Ihemedu-Steinke, Stanislava Rangelova, Michael Weber, Rainer Erbach, Gerrit Meixner, and Nicola Marsden. Simulation sickness related to virtual reality driving simulation. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 521–532. Springer, Springer International Publishing, 2017.

[101] SAE International. J2944 202302: Operational definitions of driving performance measures and statistics, Feb 2023. URL `https://www.sae.org/standards/content/j2944_202302/`.

[102] Robert JK Jacob, Audrey Girouard, Leanne M Hirshfield, Michael S Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 201–210. ACM, 2008.

[103] JiHyun Jeong. Leveraging social interactions in human-agent decision-making. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1816–1817, 2021.

[104] JiHyun Jeong, David Goedicke, Qian Yang, Wendy Ju, and Guy Hoffman. A pedestrian, a driver, and an autonomous vehicle passenger encounter an ambiguous right-of-way at the intersection, forthcoming.

[105] Wendy Ju. *The design of implicit interactions*. Springer Nature, 2022.

[106] Oskar Juhlin. Traffic behaviour as social interaction-implications for the de-

sign of artificial drivers. In *Proceedings of the 6th World Congress on Intelligent Transport Systems*, Toronto, Canada, 11 1999.

[107] Nico Kaptein, Jan Theeuwes, and Richard Van Der Horst. Driving simulator validity: Some considerations. *Transportation Research Record: Journal of the Transportation Research Board*, 1550:30–36, 1996.

[108] Joseph K Kearney, David A Noyce, Kelvin R Santiago-Chaparro, Soumyajit Chakraborty, Yuanyuan Jiang, et al. Multi-modal distributed simulation combining cars, bicyclists, and pedestrians. Technical report, University of Iowa;University of Wisconsin–Madison. Dept. of Civil and Environmental Engineering;California State University, San Marcos, 2018.

[109] Andras Kemeny. From driving simulation to virtual reality. In *Proceedings of the 2014 Virtual Reality International Conference*, page 32. ACM, 2014.

[110] Andras Kemeny and Francesco Panerai. Evaluating perception in driving simulation experiments. *Trends in cognitive sciences*, 7(1):31–37, 2003.

[111] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3): 203–220, 1993.

[112] Dagmar Kern and Albrecht Schmidt. Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '09, page 3–10, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585710. doi: 10.1145/1620509.1620511. URL `https://doi.org/10.1145/1620509.1620511`.

[113] Esko Keskinen, Mika Hatakka, Sirkku Laapotti, Ari Katila, and M Peräaho. Driver behaviour as a hierarchical system. *Traffic and Transport Psychology*, pages 9–29, 2004.

[114] Eunjee Kim and Gwanseob Shin. User discomfort while using a virtual reality headset as a personal viewing system for text-intensive office tasks. *Ergonomics*, 64(7):1–9, 2020.

[115] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023.

[116] Lars Knaack, Ann-Karolin Lache, Oliver Preikszas, Sascha Reinhold, and Michael Teistler. Improving readability of text in realistic virtual reality scenarios: Visual magnification without restricting user interactions. In *Proceedings of Mensch Und Computer 2019*, MuC'19, page 749–753, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450371988. doi: 10.1145/3340764.3344902. URL `https://doi.org/10.1145/3340764.3344902`.

[117] Christopher Knauf. 6dof rotary stewart motion simulator platform, Nov 2022. URL `https://github.com/knaufinator/6DOF-Rotary-Stewart-Motion-Simulator`.

[118] Kenji Koide, Jun Miura, and Emanuele Menegatti. A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement. *International Journal of Advanced Robotic Systems*, 16, 02 2019. doi: 10.1177/1729881419841532.

[119] Kia Kokalitcheva. People cause most self-driving car accidents in california, Aug 2018. URL `https://www.axios.com/california-people-cause-most-autonomous-vehicle-accidents-dc962265-c9bb-4b00-ae97-50427f6bc936.html`.

[120] Maxim Kolesnikov, Milos Zefran, Arnold D Steinberg, and Philip G Bashook. Periosim: Haptic virtual reality simulator for sensorimotor skill acquisition in dentistry. In *2009 IEEE International Conference on Robotics and Automation*, pages 689–694. IEEE, 2009.

[121] Sven Krome, Eric Deng, David Goedicke, Wendy Ju, Ignacio Alverez, Jaka Sodnik, Andrew Veit, and Francesco Grani. Simulator showdown: pitch your virtual ride. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, pages 25–31, 2019.

[122] Markus Kuderer, Shilpa Gulati, and Wolfram Burgard. Learning driving styles for autonomous vehicles from demonstration. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2641–2646, 2015.

[123] Timo Lajunen and Heikki Summala. Driving experience, personality, and skill and safety-motive dimensions in drivers' self-assessments. *Personality and Individual Differences*, 19(3):307–318, 1995.

[124] Timo Lajunen, Dianne Parker, and Stephen G Stradling. Dimensions of driver anger, aggressive and highway code violations and their mediation by safety orientation in uk drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 1(2):107–121, 1998.

[125] Brian Lathrop, Hua Cheng, Fuliang Weng, Rohit Mishra, Joyce Chen, Harry

Bratt, Lawrence Cavedon, Carsten Bergmann, Tess Hand-Bender, Heather Pon-Barry, et al. A wizard of oz framework for collecting spoken human-computer dialogs: An experiment procedure for the design and testing of natural language in-vehicle technology systems. In *Proc. ITS*, 2004.

[126] Hedda Lausberg and Han Sloetjes. Coding gestural behavior with the neuroges-elan system. *Behavior research methods*, 41(3):841–849, 2009.

[127] Hoe C Lee, Don Cameron, and Andy H Lee. Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accident Analysis & Prevention*, 35(5):797–803, 2003.

[128] John Lee, Nic Ward, Erwin Boer, Timothy Brown, Stacy Balk, and Omar Ahmad. Making driving simulators more useful for behavioral research – simulator characteristics comparison and model-based transformation. Technical Report N2013-016, University of Iowa, 10 2013.

[129] Suzanne E Lee, Erik CB Olsen, Walter W Wierwille, et al. A comprehensive examination of naturalistic lane-changes. Technical report, United States. National Highway Traffic Safety Administration, 2004.

[130] Lingyi Li, Shun Tian, Wencai Zhou, and Fenghui Wang. Application of driving simulators in the validation test for vehicle active safety system. *International Journal of Crashworthiness*, 28(2):159–169, 2023.

[131] Anthony Best Dynamics Limited. avds advanced vehicle driving simulator, Feb 2023. URL `https://www.abdynamics.com/en/products/simulation/avds`.

[132] Oliver Liu, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. Understanding human-robot interaction in virtual reality. In *2017 26th IEEE international*

symposium on robot and human interactive communication (RO-MAN), pages 751–757. IEEE, 2017.

[133] Ruilin Liu, Daehan Kwak, Srinivas Devarakonda, Kostas Bekris, and Liviu Iftode. Investigating remote driving over the lte network. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '17, page 264–269, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450351508. doi: 10.1145/3122986.3123008. URL `https://doi.org/10.1145/3122986.3123008`.

[134] Thomas Lopez, Olivier Dumas, Fabien Danieau, Bertrand Leroy, Nicolas Mollet, and Jean-François Vial. A playback tool for reviewing vr experiences. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pages 1–2, 2017.

[135] Katerina Mania, Bernard D. Adelstein, Stephen R. Ellis, and Michael I. Hill. Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, APGV '04, page 39–47, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581139144. doi: 10.1145/1012551.1012559. URL `https://doi.org/10.1145/1012551.1012559`.

[136] Gustav Markkula, Ruth Madigan, Dimitris Nathanael, Evangelia Portouli, Yee Mun Lee, André Dietrich, Jac Billington, Anna Schieben, and Natasha Merat. Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science*, 21(6):728–752, 2020.

[137] Nikolas Martelaro and Wendy Ju. Woz way: Enabling real-time remote interaction prototyping & observation in on-road vehicles. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 21–24. ACM, 2017.

[138] Nikolas Martelaro and Wendy Ju. WoZ Way: Enabling Real-time Remote Interaction Prototyping & Observation in On-road Vehicles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 169–182, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4335-0. doi: 10.1145/2998181.2998293. URL `http://doi.acm.org/10.1145/2998181.2998293`.

[139] Nikolas Martelaro, David Sirkin, and Wendy Ju. DAZE: A Real-time Situation Awareness Measurement Tool for Driving. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 158–163. ACM, 2015. URL `http://dl.acm.org/citation.cfm?id=2809753`.

[140] Daniel R Mayhew, Herb M Simpson, Katherine M Wood, Lawrence Lonero, Kathryn M Clinton, and Amanda G Johnson. On-road and simulated driving: Concurrent and discriminant validation. *Journal of safety research*, 42(4): 267–275, 2011.

[141] Mark McGill, Alexander Ng, and Stephen Brewster. I Am The Passenger: How Visual Motion Cues Can Influence Sickness For In-Car VR. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 5655–5668, 2017. doi: 10.1145/3025453.3026046. URL `http://eprints.gla.ac.uk/133167/%0Ahttp://dl.acm.org/citation.cfm?doid=3025453.3026046`.

[142] Mark McGill, Graham Wilson, Daniel Medeiros, and Stephen Anthony Brewster. 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*, UIST '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393201. doi: 10.1145/3526113.3545657. URL https://doi.org/10.1145/3526113.3545657.

[143] A James McKnight and Bert B Adams. Driver education task analysis. volume i: Task descriptions. final report (august 1969-july 1970). Technical report, National Technical Information Service, 1970.

[144] Natasha Merat. University of leeds driving simulator, July 2023. URL https://uolds.leeds.ac.uk/facility/driving-simulator/.

[145] Natasha Merat, A Hamish Jamson, Frank CH Lai, Michael Daly, and Oliver MJ Carsten. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, 27:274–282, 2014.

[146] John A. Michon. *A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?*, pages 485–524. Springer US, Boston, MA, 1985. ISBN 978-1-4613-2173-6. doi: 10.1007/978-1-4613-2173-6_19. URL https://doi.org/10.1007/978-1-4613-2173-6_19.

[147] David Miller, Annabel Sun, Mishel Johns, Hillary Ive, David Sirkin, Sudipto Aich, and Wendy Ju. Distraction becomes engagement in automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 1676–1680. SAGE Publications Sage CA: Los Angeles, CA, 2015.

[148] Evan Miller, Bee Hanson, Diana Hanson, and Yashar Bahman. Moving platform mode on low dynamic motion moving platforms, 2021. URL `https://docs.microsoft.com/en-us/hololens/hololens2-moving-platform`.

[149] Lanssie Mingyue Ma, Terrence Fong, Mark J. Micire, Yun Kyung Kim, and Karen Feigh. Human-robot teaming: Concepts and components for design. In Marco Hutter and Roland Siegwart, editors, *Field and Service Robotics*, pages 649–663, Cham, 2018. Springer International Publishing. ISBN 978-3-319-67361-5.

[150] Rohit Mishra, Elizabeth Shriberg, Sandra Upson, Joyce Chen, Fuliang Weng, Stanley Peters, Lawrence Cavedon, John Niekrasz, Hua Cheng, and Harry Bratt. A wizard of oz framework for collecting spoken human-computer dialogs. In *Eighth International Conference on Spoken Language Processing*, 2004.

[151] Brian Ka-Jun Mok, David Sirkin, Srinath Sibi, David Bryan Miller, and Wendy Ju. Understanding driver-automated vehicle interactions through wizard of oz design improvisation. In *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 386–392, 2015.

[152] Dylan Moore and Wendy Ju. Sound as implicit influence on human-robot interactions. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, page 311–312, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356152. doi: 10.1145/3173386.3176918. URL `https://doi.org/10.1145/3173386.3176918`.

[153] Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. Mak-

ing Noise Intentional: A Study of Servo Sound Perception. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 12–21. ACM, 2017. URL `http://dl.acm.org/citation.cfm?id=3020238`.

[154] Dominik Mühlbacher, Jutta Zimmer, Florian Fischer, and Hans-Peter Krüger. The multi-driver simulator–a new concept of driving simulation for the analysis of interactions between several drivers. *Human Centred Automation*, pages 147–158, 2011.

[155] Nadia Mullen, Judith Charlton, Anna Devlin, and Michel Bedard. Simulator validity: Behaviours observed on the simulator and on the road. In *Handbook of driving simulation for engineering, medicine and psychology*, pages 1–18. CRC Press, 2011.

[156] Takahiko Murano, Takashi Yonekawa, Masami Aga, and Sueharu Nagiri. Development of high-performance driving simulator. *SAE International Journal of Passenger Cars-Mechanical Systems*, 2(2009-01-0450):661–669, 2009.

[157] Okihiko Nakayama, Tohru Futami, Tomokazu Nakamura, and Erwin R Boer. Development of a steering entropy method for evaluating driver workload. *SAE transactions*, pages 1686–1695, 1999.

[158] Frederik Naujoks, Sebastian Hergeth, Katharina Wiedemann, Nadja Schömig, Yannick Forster, and Andreas Keinath. Test procedure for evaluating the human–machine interface of vehicles with automated driving systems. *Traffic injury prevention*, 20(sup1):S146–S151, 2019.

[159] Stefan Neumeier, Nicolas Gay, Clemens Dannheim, and Christian Facchi. On the way to autonomous vehicles teleoperated driving. In *AmE 2018-*

Automotive meets Electronics; 9th GMM-Symposium, pages 1–6. VDE, VDE, 2018.

[160] Stefan Neumeier, Philipp Wintersberger, Anna-Katharina Frison, Armin Becher, Christian Facchi, and Andreas Riener. Teleoperation: The holy grail to solve problems of automated driving? sure, but latency matters. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '19, page 186–197, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368841. doi: 10.1145/3342197.3344534. URL `https://doi.org/10.1145/3342197.3344534`.

[161] Daniel Normark. *Enacting Mobility. Studies into the Nature of Road-related Social Interaction*. PhD thesis, University of Gothenburg, 2006.

[162] Katharina Oeltze and Caroline Schießl. Benefits and challenges of multi-driver simulator studies. *IET intelligent transport systems*, 9(6):618–625, 2015.

[163] Society of Automotive Engineers. Operational definitions of driving performance measures and statistics, jun 2015.

[164] Cynthia Owsley and Gerald McGwin Jr. Vision impairment and driving. *Survey of ophthalmology*, 43(6):535–550, 1999.

[165] Türker Özkan, Timo Lajunen, Joannes El Chliaoutakis, Dianne Parker, and Heikki Summala. Cross-cultural differences in driving skills: A comparison of six countries. *Accident Analysis & Prevention*, 38(5):1011–1018, 2006.

[166] Lothar Pantel and Lars C. Wolf. On the impact of delay on real-time multiplayer games. In *Proceedings of the 12th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, NOSSDAV '02,

page 23–29, New York, NY, USA, 2002. Association for Computing Machinery. ISBN 1581135122. doi: 10.1145/507670.507674. URL https://doi.org/10.1145/507670.507674.

[167] Pablo E Paredes, Stephanie Balters, Kyle Qian, Elizabeth L Murnane, Francisco Ordóñez, Wendy Ju, and James A Landay. Driving with the fishes: Towards calming and mindful virtual reality experiences for the car. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2 (4):1–21, 2018.

[168] Martin Placek. Autonomous vehicle market size worldwide 2021-2030, Jan 2023. URL https://www.statista.com/statistics/1224515/av-market-size-worldwide-forecast/.

[169] Victor Ponce and Bessam Abdulrazak. Context-aware end-user development review. *Applied Sciences*, 12(1), 2022. ISSN 2076-3417. doi: 10.3390/app12010479. URL https://www.mdpi.com/2076-3417/12/1/479.

[170] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, number 3.2, page 5. Kobe, Japan, 2009.

[171] J. Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):257–266, 1983.

[172] James Reason, Antony Manstead, Stephen Stradling, James Baxter, and Karen Campbell. Errors and violations on the roads: a real distinction? *Ergonomics*, 33(10-11):1315–1332, 1990.

[173] Lena Rittger, Dominik Muehlbacher, Christian Maag, and Andrea Kiesel. Anger and bother experience when driving with a traffic light assistant: A multi-driver simulator study. *Proceedings of the Human Factors and Ergonomics Society Europe*, pages 41–51, 2015.

[174] SafeMotorist.com. Right of way, 2006. URL http://www.safemotorist.com/Articles/Right_of_Way.aspx.

[175] Paul Sârbescu, Predrag Stanojevic, and Dragan Jovanovic. A cross-cultural analysis of aggressive driving: evidence from serbia and romania. *Transportation research part F: traffic psychology and behaviour*, 24:210–217, 2014. ISSN 1369-8478. doi: 10.1016/j.trf.2014.04.002. URL http://dx.doi.org/10.1016/j.trf.2014.04.002.

[176] Anna Schieben, Matthias Heesen, Julian Schindler, Johann Kelsch, and Frank Flemisch. The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 43–46. ACM, 2009.

[177] G Schmidt, M Kiss, E Babbel, and A Galla. The wizard on wheels: Rapid prototyping and user testing of future driver assistance using wizard of oz technique in a vehicle. In *Proceedings of the FISITA 2008 World Automotive Congress, Munich*, 2008.

[178] Paul Schmitt, Nicholas Britten, JiHyun Jeong, Amelia Coffey, Kevin Clark, Shweta Sunil Kothawade, Elena Corina Grigore, Adam Khaw, Christopher Konopka, Linh Pham, Kim Ryan, Christopher Schmitt, and Emilio Frazzoli. Can cars gesture? a case for expressive behavior within autonomous vehicle

and pedestrian interactions. *IEEE Robotics and Automation Letters*, 7(2):1416–1423, 2022. doi: 10.1109/LRA.2021.3138161.

[179] Paul Schmitt, Nicholas Britten, JiHyun Jeong, Amelia Coffey, Kevin Clark, Shweta Sunil Kothawade, Elena Corina Grigore, Adam Khaw, Christopher Konopka, Linh Pham, Kim Ryan, Christopher Schmitt, Aryaman Pandya, and Emilio Frazzoli. nureality: A vr environment for research of pedestrian and autonomous vehicle interactions, 2022.

[180] Brandon Schoettle and Michael Sivak. A preliminary analysis of real-world crashes involving self-driving vehicles. *University of Michigan Transportation Research Institute*, 2015.

[181] David Serje and Estefany Acuña. Driving and flying simulators: a review on relevant considerations and trends. *Transportation research record*, 2676(3): 551–570, 2022.

[182] Orit Shechtman, Sherrilene Classen, Kezia Awadzi, and William Mann. Comparison of driving errors between on-the-road and simulated driving assessment: a validation study. *Traffic injury prevention*, 10(4):379–385, 2009.

[183] Gustavo Silvera, Abhijat Biswas, and Henny Admoni. Dreyevr: Democratizing virtual reality driving simulation for behavioural & interaction research. In *HRI*, pages 639–643, 2022.

[184] David Sirkin, Kerstin Fischer, Lars Jensen, and Wendy Ju. Eliciting Conversation in Robot Vehicle Interactions. In *2016 AAAI Spring Symposium Series*, March 2016. URL `http://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12755`.

[185] SKARREDGHOST. The metaverse enters the gartner hype cycle (but with a 10+ years outlook), Aug 2022. URL `https://skarredghost.com/2022/08/20/metaverse-gartner-hype-cycle/`.

[186] JJ Slob. State-of-the-art driving simulators, a literature survey. *DCT report*, 107, 2008.

[187] Joseph Smyth, Jonathan Robinson, Rebecca Burridge, Paul Jennings, and Roger Woodman. Towards the management and mitigation of motion sickness–an update to the field. In *Congress of the International Ergonomics Association*, pages 834–840. Springer, Springer International Publishing, 2021.

[188] Inc. StereoLabs. Meet zed 2, December 2019. URL `https://www.stereolabs.com/zed-2/`.

[189] Jonathan Stevens, Peter Kincaid, and Robert Sottilare. Visual modality research in virtual and mixed reality simulation. *The Journal of Defense Modeling and Simulation*, 12(4):519–537, 2015.

[190] Jack Stewart. Humans just can't stop rear-ending self-driving cars-let's figure out why, Oct 2018. URL `https://www.wired.com/story/self-driving-car-crashes-rear-endings-why-charts-statistics/`.

[191] Matúš Šucha. Road users' strategies and communication: driver-pedestrian interaction. *Transport Research Arena (TRA)*, 2014.

[192] Mark JM Sullman, M Eugenia Gras, Monica Cunill, Monserrat Planes, and Silvia Font-Mayolas. Driving anger in spain. *Personality and Individual Differences*, 42(4):701–713, 2007.

[193] Daniel Szafir. Mediating human-robot interactions with virtual, augmented, and mixed reality. In *International Conference on Human-Computer Interaction*, pages 124–149. Springer, 2019.

[194] Seyyed Meisam Taheri, Kojiro Matsushita, Minoru Sasaki, et al. Virtual reality driving simulation for measuring driver behavior and characteristics. *Journal of transportation technologies*, 7(02):123, 2017.

[195] Leila Takayama and Caroline Pantofaru. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5495–5502. IEEE, 2009.

[196] Unity Technologies. Unity game engine, Jul 2023. URL `https://unity.com`.

[197] Jack Tilley. Grand theft auto v: The rise and fall of the diy self-driving car lab. *Forbes*, Oct 2017. URL `https://www.forbes.com/sites/aarontilley/2017/10/04/grand-theft-auto-v-the-rise-and-fall-of-the-diy-self-driving-car-lab/?sh=5b0cd1417d7a`.

[198] Kalle Toiskallio. The impersonal flâneur: navigation styles of social agents in urban traffic. *Space and Culture*, 5(2):169–184, 2002.

[199] M. Tonnis, V. Broy, and G. Klinker. A survey of challenges related to the design of 3d user interfaces for car drivers. In *3D User Interfaces (3DUI'06)*, pages 127–134. IEEE Computer Society, 2006. doi: 10.1109/VR.2006.19.

[200] UltraLeap. Leap Motion Orion SDK, 2016. URL `https://developer.leapmotion.com/orion`.

[201] Unity Technologies. Unity 3d, May 2019. URL `https://unity3d.com/unity/whats-new/2018.4.0`.

[202] Valve. Steam, 2003. URL https://store.steampowered.com/app/250820/SteamVR/.

[203] Valve Software. Openvr, May 2019. URL https://github.com/ValveSoftware/openvr.

[204] Jan BF Van Erp and Pieter Padmos. Image parameters for driving with indirect viewing systems. *Ergonomics*, 46(15):1471–1499, 2003.

[205] Erik Vinkhuyzen and Melissa Cefkin. Developing socially acceptable autonomous vehicles. In *Ethnographic Praxis in Industry Conference Proceedings*, number 1, pages 522–534. Wiley Online Library, 2016.

[206] Marcel Walch, Julian Frommel, Katja Rogers, Felix Schüssel, Philipp Hock, David Dobbelstein, and Michael Weber. Evaluating vr driving simulation from a player experience perspective. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2982–2989, 2017.

[207] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, pages 27–35, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4491-3. doi: 10.1145/2993369.2993381. URL http://doi.acm.org/10.1145/2993369.2993381.

[208] Peter Wang, Srinath Sibi, Brian Mok, and Wendy Ju. Marionette: Enabling on-road wizard-of-oz autonomous driving studies. In *Proceedings of the 2017*

ACM/IEEE International Conference on Human-Robot Interaction, pages 234–243. ACM, 2017.

[209] Ginger S. Watson and Yiannis E. Papelis. The iowa driving simulator: Using simulation for human performance measurement. In Sydne J. Carlson-Newberry and Rebecca B. Costello, editors, *Emerging Technologies for Nutrition Research: Potential for Assessing Military Performance Capability*, chapter 26, pages 551–568. Institute of Medicine (US) Committee on Military Nutrition Research, Oxford, 1997.

[210] Gary L Wells and Paul D Windschitl. Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9): 1115–1125, 1999.

[211] Wikipedia. Yield sign, 2019. URL `https://en.wikipedia.org/wiki/Yield_sign#cite_note-14`.

[212] Tom Williams, Daniel Szafir, Tathagata Chakraborti, Ong Soh Khim, Eric Rosen, Serena Booth, and Thomas Groechel. Virtual, augmented, and mixed reality for human-robot interaction (vam-hri). In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 663–664, 2020.

[213] Yan Yan, Ke Chen, Yu Xie, Yiming Song, and Yonghong Liu. The effects of weight on comfort of virtual reality devices. In *International Conference on Applied Human Factors and Ergonomics*, pages 239–248. Springer, Springer International Publishing, 2018.

[214] Steven Yang, Brian Mok, David Sirkin, Hilliary Ive, Rohan Maheshwari, Kerstin Fischer, and Wendy Ju. Adventures of an Adolescent Trash Barrel: Ex-

periences Developing Socially Acceptable Interactions for Everyday Robots. In *ROMAN '15*, Kobe, Japan, mar 2015. IEEE.

[215] Sharon Yavo-Ayalon, Swapna Joshi, Yuzhen (Adam) Zhang, Ruixiang (Albert) Han, Narges Mahyar, and Wendy Ju. Building community resiliency through immersive communal extended reality (cxr). *Multimodal Technologies and Interaction*, 7(5), 2023. ISSN 2414-4088. doi: 10.3390/mti7050043. URL `https://www.mdpi.com/2414-4088/7/5/43`.

[216] Dohyeon Yeo, Gwangbin Kim, and SeungJun Kim. Maxim: Mixed-reality automotive driving ximulation. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 460–464. IEEE, 2019. doi: 10.1109/ISMAR-Adjunct.2019.00124.

[217] David M Zaidel. A modeling perspective on the culture of driving. *Accident Analysis & Prevention*, 24(6):585–597, 1992.

[218] Maryum Zaigham, Jakaria Dasan, and Christie Pei-Yee Chin. A systematic literature review on ridesharing: Determinants, theoretical grounds and methodologies. *Management*, 7(27):187–200, Sep 2022.

[219] Hangyu Zhou, Yuichiro Fujimoto, Masayuki Kanbara, and Hirokazu Kato. Effectiveness of virtual reality playback in public speaking training. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 462–466, 2020.

[220] Xin Zou, Steve O'Hern, Barrett Ens, Selby Coxon, Pascal Mater, Raymond Chow, Michael Neylan, and Hai L. Vu. On-road virtual reality autonomous vehicle (vrav) simulator: An empirical study on user experience. *Transportation Research Part C: Emerging Technologies*, 126:103090, 2021. ISSN 0968-

090X. doi: https://doi.org/10.1016/j.trc.2021.103090. URL https://www.sciencedirect.com/science/article/pii/S0968090X2100111X.

APPENDIX A

APPENDIX

A.1 External Code Repositories

VR-OOM `https://github.com/FAR-Lab/vr-oom`

XR-OOM `https://github.com/FAR-Lab/XR-OOM`

***XR-OOM*- localization** `https://github.com/FAR-Lab/xr-oom-localization`

Checklist template `https://github.com/DavidGoedicke/SimpleToDoList`

StrangeLand `https://github.com/FAR-Lab/StrangeLand`

ReRun `https://github.com/FAR-Lab/Rerun`

***StrangeLand* Data Analysis** `https://github.com/FAR-Lab/XCulturalDataAnalysis`

***StrangeLand* Visualization** `https://github.com/FAR-Lab/VideoVisualizationXC`

A.2 *XR-OOM* Appendix

A.2.1 System Implementation Details

In this appendix section, we wanted to share the specific technical details for our prototype implementation, to make replication easier.

Rendering computer The computational requirements for the graphics rendering are driven by the requirements from the Varjo XR-1 headset[1]. The rendering computer also handles the tracking of the participant's gaze within the mixed reality environment, and recording of the headsets video stream. Our setup uses a

[1]`https://varjo.com/use-center/get-started/system-requirements/headsets/xr-1-vr/`

custom-built desktop tower computer; the essential components are: Mainboard: Gigabyte Z390 Designare, CPU: i9 9900K, RAM: 2×16GB 3200 DDR4, GPU: NVIDA 2080 Ti, Power supply: SeaSonic Focus Plus 750 W Platinum.

While a laptop computer would be preferable to a desktop computer for this system, at submission time, only one laptop was available that would meet the computational demands of the *XR-OOM* system, and all such platforms were in short supply due to the pandemic-related worldwide chip shortage. It is not clear that the one laptop that meets the computational needs would be able to run within the thermal limits of the vehicle, because vehicles are significantly warmer than the information work environments laptops are generally designed for.

Networking The SmartTrack3 is connected to the graphics rendering computer via a 5-port Ethernet Switch (NETGEAR GS105GE). Additionally, a Netgear M1 router provided internet access (used to upload experiment data), and to act as the DHCP server.

Power System The computer, headset (Varjo XR-1), networking, and in-car localization (SmartTrack3) have a total rated power draw of about 850 W; this is more than is available from most vehicles via the 12 V power ports. In our set-up, the *XR-OOM* system is powered by a 1.5 kW modified sine-wave power supply that is supplied by three 12 V deep-cycle batteries. Additionally, a GoalZero 400 that was continuously charged from the Cars 12V outlet was providing power some of the hardware, in a rudimentary attempt to load balance the power requirement.

LIDAR A OS1-64 2nd Generation LIDAR sensor from Ouster was mounted at the front of the vehicle, running at 1024 beams with 20 HZ. This sensor was used to build a depth map of the environment around the research vehicle.

IMU An Xsense MTi-300 IMU was mounted in the center of the vehicle, and connected over USB to the ROS core computer; this enabled tracking of the orientation of the vehicle in space.

Odometry Odometry data from our research vehicle, a 2015 Toyota PriusV, was obtained from the vehicle's CAN Bus using a Korlan USB2CAN module. This was primarily used to track the forward velocity of the vehicle; data was sent over USB to the ROS core computer.

The pace of hardware development and Moore's Law predicts that the system performance specifications here will likely be more easily achieved using consumer-grade equipment in the hopefully non-distant future.

A.2.2 Self-reported Measures

We also recorded results from each participant rating their own assessments for each condition.

Following each condition's activity, participants were asked to respond to four questions on a 5-point Likert scale to assess their impressions of ease of use, comfort, control, and performance. The responses were reverse-scored where necessary so that higher values (5) mean a more favorable response (i.e., more comfortable, easier driving, more control, better performance) and lower values (1) mean a less favorable response. (See Appendix for the questions sheets)

The responses to these four questions generally agree and are highly correlated. Comparing driving condition A (no headset) to conditions B and C (with the headset), we see condition A was rated, on average, 3.7 for ease of use while the headset conditions were rated 2.85. Similarly, for comfort, condition A was rated 4.3 while

Prompt	**Condition A**	**Condition B**	**Condition C**
Ease of Use	3.7	3.3	2.4
Comfort	4.3	3.2	2.6
Control	4.3	3.9	3.3
Self-Reported Performance	3.9	3.0	2.8

Table A.1: Average ratings (out of 5) across participants per driving condition, where 1 was least favorable and 5 was most favorable

the headset conditions were rated 2.9. For control, condition A was rated 4.3 and the headset conditions were rated 3.6. Finally, for self-rated performance, condition A was rated 3.9 and the headset conditions were rated 2.9. This shows that, participants found the headset more difficult to drive with and rated it a full point lower or more in every category. Similarly, if we compare the two headset conditions B (real cones) and C (virtual cones), we see that, on average, condition B was rated 3.3 for ease of use while condition C was rated 2.4. For comfort, condition B was rated 3.2 and condition C was rated 2.6. For control, condition B was rated 3.9 and condition C was rated 3.3. And for self-reported performance, condition B was rated 3.0 and condition C was rated 2.8. Thus, participants not only found the headset conditions more difficult, they also generally found the virtual cones more difficult than the real cones (again often with a full point difference in rating). The one disagreement amongst the categories here is that the participants appeared to believe they performed similarly in conditions B and C even though they found condition C to be less easy to use, less comfortable, and felt they had less control.

The distribution of these responses by condition are shown in Figure A.1. Condition C (with VR Headset and virtual cones) was more consistently rated as less easy and less comfortable. Whereas condition A (normal driving condition without headset) was rated consistently with a high self-reported performance.

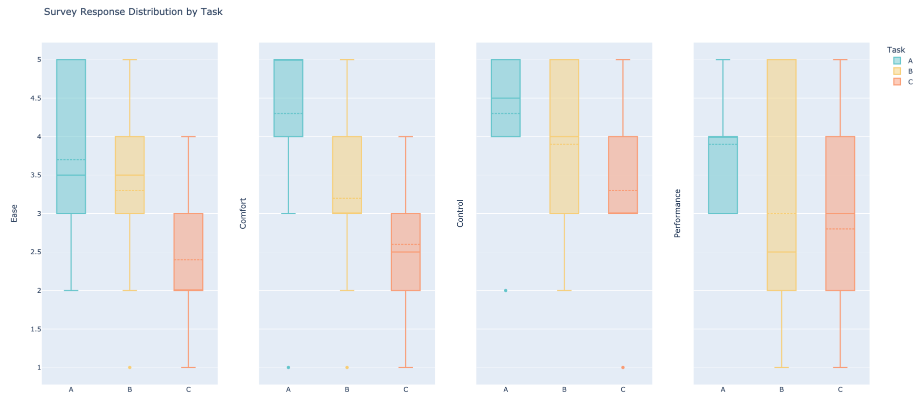


Figure A.1: Distribution of responses to survey prompts by task (dotted line = mean, solid line = median)

An exploratory factor analysis (EFA) was run on the responses to determine inter-item agreement. The EFA identifies a single factor forms across all four questions where none of the questions are unique (all uniqueness scores ¡ 0.48). Inter-question correlations, as mentioned earlier, are quite high. This indicates that these dimensions generally agree. We adjusted the ordering of the conditions in order to counterbalance our results and remove ordering effects. The final average participant responses for each condition by prompt is shown in Table A.1. Again, this is on a Likert scale from 1 to 5 where 1 is a less favorable response and 5 is a more favorable response. This clearly shows that condition favorability is consistently ranked A, B, and C from most favorable to least favorable across the four metrics where the widest range is on comfort.

Besides the Likert scale questions, we also asked the participant an additional four open-ended questions after each scenario. These questions attempted to identify difficulties and the general impression the participant had while completing the driving tasks.

The most relevant results are from the questions Q1: *Was there anything that felt outside of your control?*, Q2:*Was there a particularly hard task?*, and Q3:*Was there a*

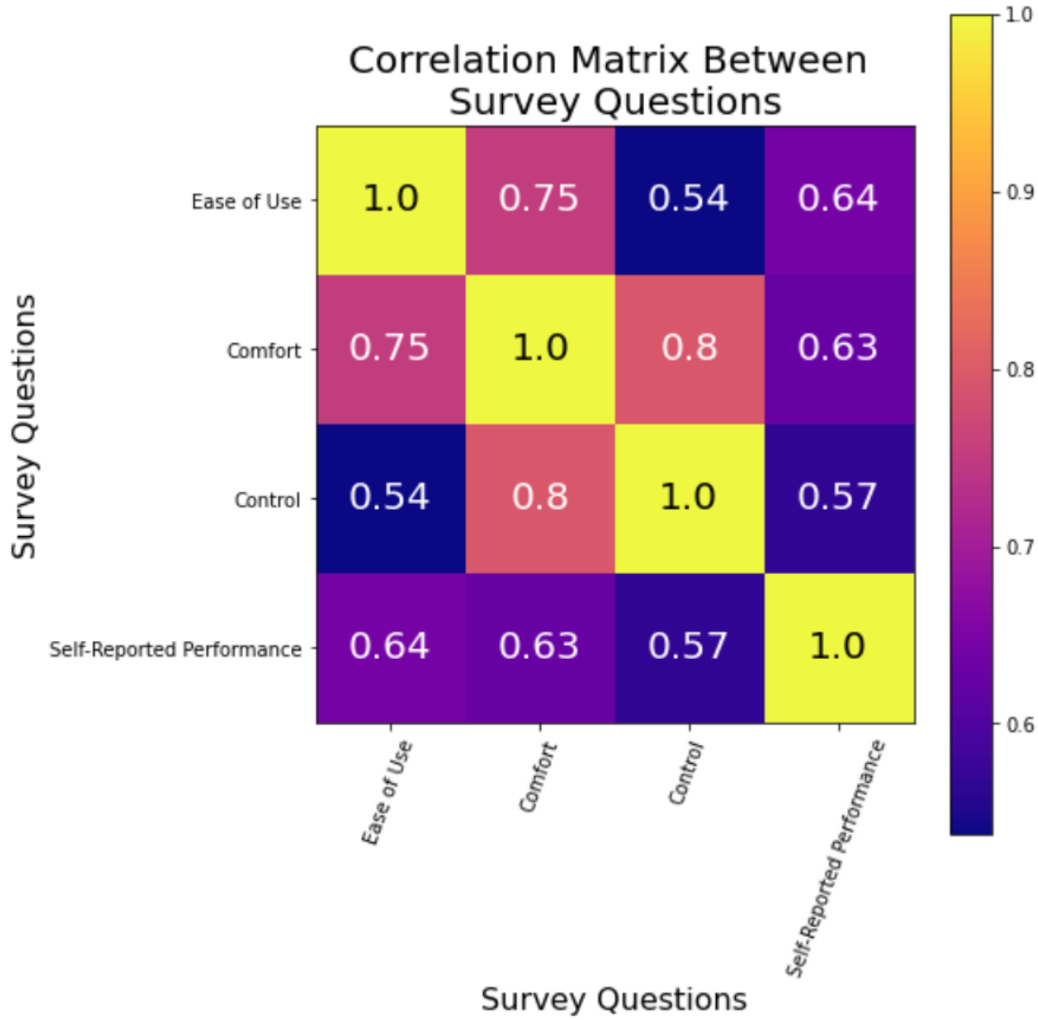


Figure A.2: Correlation matrix for the four survey questions

particularly easy task?.

Q1: For conditions A and B the most frequent response was a sense of unfamiliarity with the vehicle. For Condition C most participants mention the virtual cones as an aspect that was outside of their control. The cones could sometimes move somewhat erratically if the participant made fast head movements.

Q2: The slalom task(step2) was clearly the hardest task for people to complete with the most mistakes. This was backed up by the self report results where many participants mentioned “Step 2” being the hardest independent of the condition.

Q4: Participants report that the stop line task (Step 4) was the easiest to complete.

A.2.3 Standard Operating Procedure

In the development of *XR-OOM*, a comprehensive Standard Operating Procedure (SOP) was meticulously crafted. This SOP was essential for several reasons. First and foremost, it prioritized the safety of participants and researchers, outlining specific measures to avoid COVID infections and ensure a secure testing environment. Additionally, the detailed SOP facilitated the replication and accessibility of *XR-OOM* to other researchers, enabling them to adopt and utilize the system effectively for their own studies.

1

EXPERIMENT PREPARATION

Goal Zero .... TURN-ON OUTPUT & 12V INPUT
360°GoPro .... POWER-ON & VERIFY 5V DC
Outside GoPro .... POWER-ON & VERIFY 5V DC
LIDAR .... MOUNT
AR-Computer .... (RE)BOOT
- Plugin Power
- Turn on 1500W DC/AC

ART Tracker .... START & VERIFY
XR-1 .... START & VERIFY
- Video pass through
- No Screen Tearing
- Low frame-delay
- Pose Tracking

ROS Core .... (RE)BOOT
- Run Setup **./SetupExperiment.sh**
- LIDAR - verify
- IMU - verify
- CanBus - verify

Vehicle ....
- Drive Vehicle to starting point
- Start AC

Participant Items ....
- Varjo XR-1
- Cleaned Headset Face covers
- Participant Instructions
- Alcohol wipes

2

PARTICIPANT INTRODUCTION

Read Introduction ....

- Thank you for coming! I am [Researcher's Name] working for Cornell and I'll be supervising your driving experiment for the next 30 to 40 minutes. Before we get started, a couple of details about the study.
- There are 2 different tasks, cockpit and driving tasks. You will perform both of them three times. Twice with, and once without the MR-Headset.
- This is a test of the system not of you. If its hard for you it means the system needs to be better.
- Do not do anything that makes you feel unsafe and out of control.
- Please note that you can stop the experiment at anytime for any reason. Please let us know if you feel motion sick or unwell in any other way.

- Drive Slowly!
- We do not expect the visual system to fail, but if it does, just stop the vehicle by stepping on the brake.
- Please review and follow the hygiene standards.

Show Participant .... GLOVES & ALCOHOL WIPES

Figure A.3: First two pages of the SOP to run the the *XR-OOM* simulator.

3

EXPERIMENT PROCEDURE

Experiment Preparation . . . .
Participant Introduction . . . .
Condition N (No headset) . . . .

- Complete Cockpit Tasks
- Complete Driving Tasks 1-5
- Condition Questionnaire

Condition with (Headset) . . . .

- Explain: How to put on and off the headset.
- Verify MR headset functionality with participant
- Complete Cockpit Tasks
- Complete Driving Tasks 1-5
- Condition Questionnaire

Post Experiment Questionnaire . . . .
Close Experiment . . . .
Store Data . . . . `COMBINE.TECH.CORNELL.EDU:5000`

7

ACCIDENT CHECKLIST

- Stop the study!
- Check for injury and triage any medical situation.
- In case of serious injuries, call 911
- Inform the owner if we hit any parked car or traffic device
- Call the police and report damage if there is significant damage: (212) 832-4545
- Exchange contact and insurance information with other parties
- Note traffic and weather conditions and sketch a diagram of the collision scene.
- Alert the University and associated insurance within 24 hours.
- Alert IRB that there was an adverse event, and not continue research until we got further permission.

The next pages are for the participant

Figure A.4: Last two pages of the SOP to run the the *XR-OOM* simulator.

A.2.4 Cockpit Tasks

8

1 HYGIENE STANDARD FOR RESEARCH VEHICLE USE

Participants, research and safety driver should always wear a mask.

- Cleaning wipes, and gloves need to be accessible to the participant.
- Extra hygiene covers for the AR headset are available.
- Before each research day, the researcher will wipe the common touch areas in the vehicle (steering wheel, door handles). The AR Headset will be wiped for every user.

COCKPIT TASKS	
Turn Vehicle On	
Adjust seat	
Put on Seat belt	
Adjust mirrors	LEFT, RIGHT & CENTER
Turn signals	LEFT & RIGHT
Headlights	TURN ON/OFF
Wipers	
Hazard lights	
Parking brake	ENGAGE & RELEASE
Verbally explain which indicator lights are visible	

Now please follow the driving instructions shown in the next few pages. There are 7 steps, all leading into each other. The black dots on the pages signify the traffic cones in front of you.

Please find your bearing before you start driving.

Figure A.5: Print-out of the cockpit tasks we asked participants to conduct. The checklists template can be found in Appendix A.1.

A.2.5 Trajectory Comparison

The driving performance was measured using post-facto analysis by computing the Fréchet distance between the normalized path driven by participants, and an ”ideal” path. These paths were confirmed against observable driving performance in the participant video recordings.

Since the tracks were normalized for the comparison, the Fréchet distance does not measure in meters. However, the range of value still represent a correct and incorrect execution of the tasks, as shown in Figure A.8. Large Fréchet distance indicates that participant's driving path deviates from the path in instruction. For example, when driving under condition C, participant 5 made a right turn earlier than instructed in step 1. Participant 5 did not repeat this mistake under condition A and completed step 1 successfully. Both paths are visualized in Figure A.7. Under condition C, participant 5's mistake leads to a Fréchet distance of 0.415, and the Fréchet distance reduces to 0.146 when participant 5 followed the instructed path successfully under condition A.

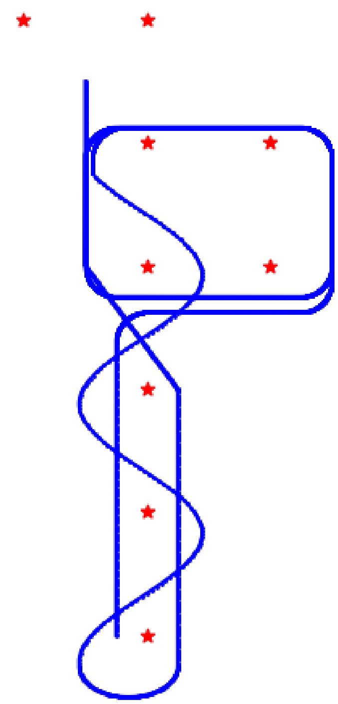


Figure A.6: Synthetic ground truth path.

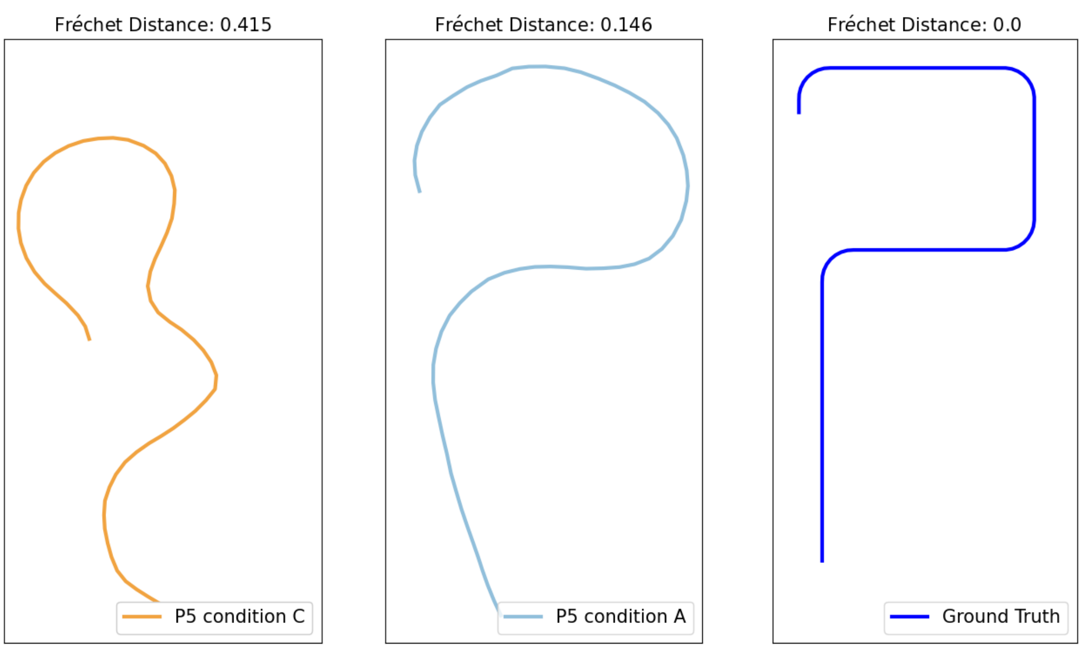


Figure A.7: Comparison between participant 5's driving performance for step 1 under two conditions.

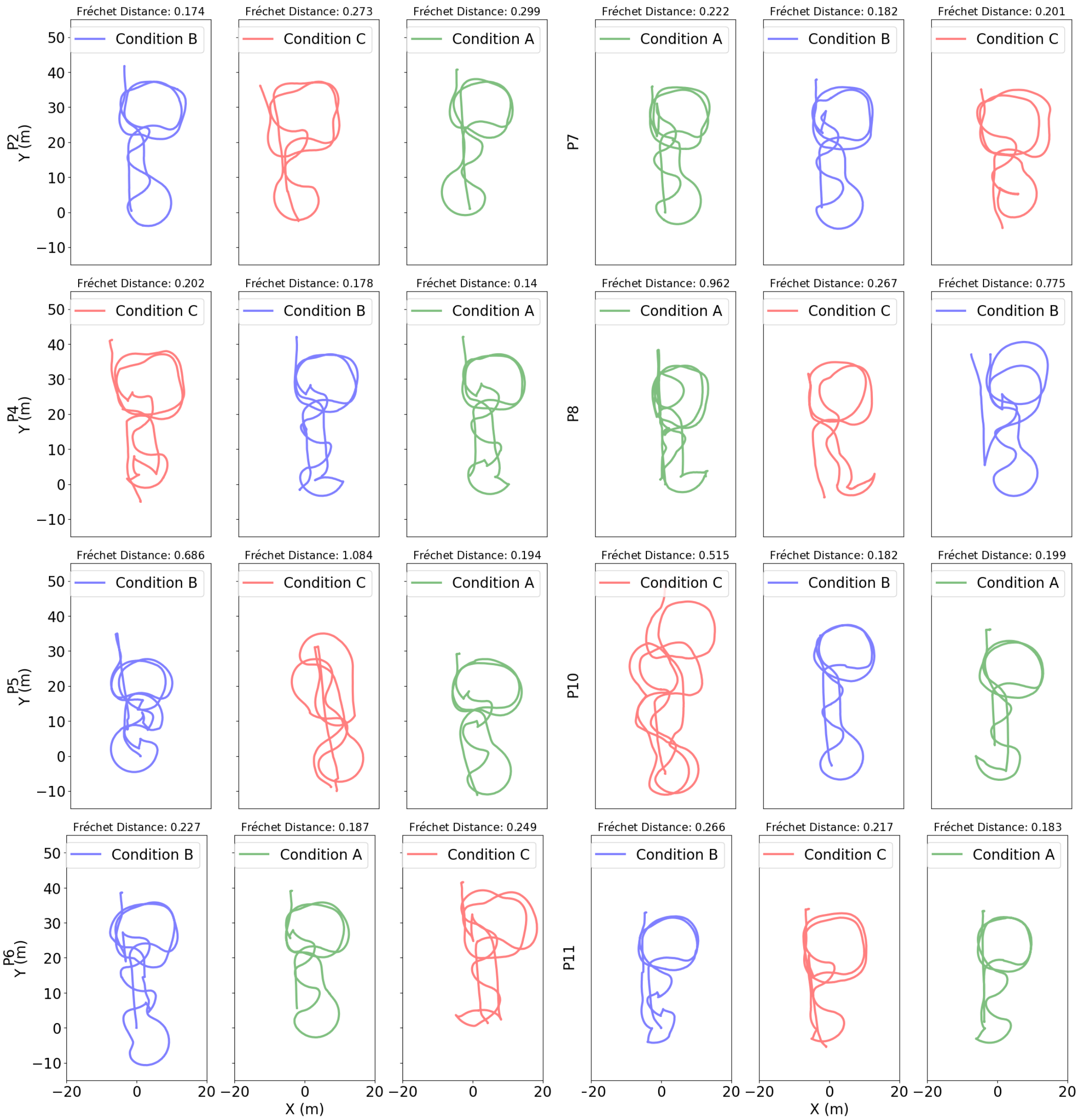


Figure A.8: Trajectories for each participant per condition.

A.3 *StrangeLand*

A.3.1 Questionnaire

These are the different questionnaires used during the study.

Certainty

After each scenario, we plan to always ask 3 questions about certainty:

I felt the other driver(s) drove well:

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

I clearly understood the intent of the other driver(s):

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

I felt confident about my own actions

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

Behaviors

We ask about these behaviors depending on the interaction that the user gets placed in. Below are the questions asking about specific behaviors.

Turn Signals

Which turn signals, if any, did you use?

- No turn signals
- Left turn signal
- Right turn signal

Which turn signals, if any, did the other car use?

- No turn signals
- Left turn signal
- Right turn signal

Did you intend for that signal to be seen by another driver in particular?

- No
- Yes
- Unsure

Did you intend for that signal to be seen by more than one driver?

- No
- Yes
- Unsure

Did you interpret the signal as meant for you in particular?

- No
- Yes
- Unsure

False Starts

Who was the first to completely cross the intersection?

- Me
- The other driver

- Unsure

At the intersection, who moved first?

- Me
- The other driver
- We moved simultaneously
- Not sure

Did you stop again after moving?

- No
- Yes
- Unsure

Did the other driver stop again after moving?

- No
- Yes
- Unsure

Did either of you stop again after moving?

- I stopped again
- The other driver stopped again
- Neither of us stopped again
- Both of us stopped again

Why did you move first?

- Because I was uncertain about whose turn it was
- Because I realized that it was my turn
- Other
- Unsure

Why did you not stop again?

- Because I was certain that it was my turn
- Because the other driver realized it was my turn
- Because the other driver gave me their turn
- Because I stole the other driver's turn
- Other
- Unsure

Why did you stop again?

- Because I was still uncertain about whose turn it was
- Because I realized it was the other driver's turn
- Because the other driver stole my turn
- Other
- Unsure

Why did the other driver move first?

- Because they were uncertain about whose turn it was
- Because it was their turn
- Other
- Unsure

Why didn't the other driver stop again?

- Because they were certain that it was their turn
- Because I realized it was their turn
- Because the other driver gave me their turn
- Because they stole my turn
- Other
- Unsure

Why did the other driver stop again?

- Because they were still uncertain about whose turn it was
- Because they realized it was my turn
- Other
- Unsure

Why did you and the other driver move simultaneously?

- Because we were both uncertain about whose turn it was
- Because we both thought it was our own turn
- Other
- I don't know why

Overtaking Who overtook the other first?

- Me
- The other driver
- Neither of us overtook the other
- Not sure

After overtaking the other driver, did you remain ahead of them?

- No
- Yes
- Unsure

Did you overtake the other driver a second time?

- No
- Yes
- Unsure

After the other driver overtook you, did they remain ahead of you?

- No

- Yes
- Unsure

Did the other driver overtake you a second time?

- No
- Yes
- Unsure

Why did you overtake the other driver?

- I didn't overtake the other driver on purpose
- The other driver was too slow
- I was going to be late for my flight
- I like to drive fast
- I like to drive faster than other people
- Other
- Unsure

Why didn't you remain ahead?

- The other driver sped up
- The other driver kept going the same speed
- I slowed down
- I kept going the same speed
- I realized that I had enough time to make my flight
- I realized that I was going too fast
- Other
- Unsure

Why did you remain ahead?

- The other driver slowed down

- The other driver kept going the same speed
- I sped up
- I kept going the same speed
- I was still worried about catching my flight
- I wanted to keep driving fast
- I wanted to keep driving faster than the other driver
- Other
- Unsure

Why did the other driver overtake you?

- The other driver didn’t overtake me on purpose
- I was driving too slow
- The other driver was worried about being somewhere on time
- The other driver likes to drive fast
- Other
- Unsure

Why didn’t the other driver remain ahead?

- I sped up
- I kept going the same speed
- The other driver slowed down
- The other driver kept going the same speed
- The other driver realized that they had enough time
- The other driver realized that they were going too fast
- Other
- Unsure

Why did the other driver remain ahead?

- I slowed down

- I kept going the same speed
- The other driver sped up
- The other driver kept going the same speed
- The other driver was still worried about being somewhere on time
- The other driver wanted to keep driving fast
- The other driver wanted to keep driving faster than other people
- Other
- Unsure

Cutting off

Did you cut off the other driver?

- No
- Yes
- Unsure

Did the other driver cut you off?

- No
- Yes
- Unsure

Did you intend to cut off the other driver?

- No
- Yes
- Unsure

Why did you cut off the other driver?

- The other driver was in my way
- So that I could make my exit

- To punish the other driver for something they did
- For fun
- To reach my destination faster
- To prevent the other driver from reaching their destination
- Other
- Unsure

Did the other driver intend to cut you off?

- No
- Yes
- Unsure

Why did the other driver cut you off?

- I was in the other driver's way
- So that they could make their exit
- To punish me for something I did
- For fun
- To reach their destination faster
- To prevent me from reaching my destination
- Other
- Unsure

Tailgating

Did you tailgate the other driver?

- No
- Yes
- Unsure

Did the other driver retaliate?

- No

- Yes
- Unsure

Did the other driver tailgate you?

- No
- Yes
- Unsure

Did you retaliate?

- No
- Yes
- Unsure

Why did you tailgate the other driver?

- To punish the other driver for something they did
- Because I was impatient
- For fun
- Other
- Unsure

Why did the other driver tailgate you?

- To punish me for something I did
- Because they were impatient
- For fun
- Other
- Unsure

Mutual Gaze

Did you and the other driver make eye contact?

- Yes
- No
- No, but I was trying to make eye contact with the other driver
- No, but the other driver was trying to make eye contact with me
- I don’t know

Who tried to first establish eye contact?

- Me
- The other driver
- Both of us
- Unclear

Why did you attempt to make eye contact with the other driver?

- To let the other driver know that I was moving
- To let the other driver know they could move
- To check the other driver’s attention
- Other

Why didn’t you attempt to make eye contact with the other driver?

- I didn’t want to validate the other driver’s behavior
- I was focused on a more important task