Strangers in a Strange Land: New Experimental System for Understanding Driving Culture Using VR

David Goedicke, *Graduate Student Member, IEEE*, Carmel Zolkov, Natalie Friedman, Talia Wise, Avi Parush, and Wendy Ju

***Abstract*—In order for autonomous vehicles 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 might help computer-driven cars drive like locals and recognize how local drivers might be signaling through hand/body movement and motion of their vehicles. To this end, we have created an experimental system 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 make sure 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.**

***Index Terms*—Human factors, vehicle safety, intelligent vehicles, road vehicles.**

I. Introduction

IN DISCUSSING Uber's driverless car experiment, Uber's Engineering Director, Raffi Krikorian, stated, "if we can drive in Pittsburgh, we can drive anywhere." [1] 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. Even people with decades of driving experience can find themselves as strangers in a strange land when they are behind the wheel away from home.

Manuscript received September 30, 2021; revised December 31, 2021; accepted February 15, 2022. Date of publication February 23, 2022; date of current version May 2, 2022. This work was supported in part by the Jacobs Ruch Award and in part by the National Science Foundation under Grant 2107111. This work was conducted under Cornell University IRB Protocol under Grant 1812008479 and Technion under Grant 56-2019 IRB Protocol. The review of this article was coordinated by Prof. Fabrizio Lamberti. *(Corresponding author: David Goedicke.)*

David Goedicke, Natalie Friedman, and Wendy Ju are with the Department of Information Science, Cornell Tech, New York, NY 10044 USA (e-mail: dg536@cornell.edu; nvf4@cornell.edu; wendyju@cornell.edu).

Carmel Zolkov, Talia Wise, and Avi Parush are with the Faculty of Industrial Engineering and Management Israel Institute of Technology Technion City, Haifa 3200003, Israel (e-mail: carmelzolkov@gmail.com; talia.wise1@gmail.com; aparush@technion.ac.il).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Cornell IRB # IRB0008479, Virtual Reality Driving Simulation. Technion IRB #56-2019, Exploring the signaling, driving behaviors and situational awareness of drivers from different cultures using a virtual reality driving simulation.

This article has supplementary downloadable material available at https://doi.org/10.1109/TVT.2022.3152611, provided by the authors.

Digital Object Identifier 10.1109/TVT.2022.3152611

As we move into an age of computer-driven vehicles, our understanding of regional differences in driving will need to develop beyond anecdote and observation towards concrete understandings of the parameters, behaviors, and interaction patterns that differentiate driving in one locale from another. Today, autonomous cars are programmed to drive within the boundaries of the law. Still, they seem to elicit higher-than-normal [2] rates of accidents because they do not conform to local driving norms [3]. Officially, many of these accidents are classified as the fault of the non-AV car [4], 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.

To do so, we need to profile differences in human driving behavior and interaction in ways that are machine-interpretable; to date, no system or method exists to make this possible.

This paper describes the design and proof-of-concept test of StrangeLand, an experimental system and method to profile behaviors and interaction drivers exhibit when interacting with each other in traffic. As part of the design effort for this project, we designed a portable setup for the driving simulation experiments, developed a multi-participant virtual reality (VR) driving simulation environment, brainstormed and tested interactive driving scenarios, instrumented the experiment to capture participant driving behavior, implicit and explicit interaction and subjective evaluation of each driving interaction, and developed an analysis tool with which researchers can replay and analyze the driving interactions. We tested this system in two locations to verify that the system actually captures differences in driving interaction between cultures. This work is the first of its kind. It makes an artifact contribution that facilitates previously impossible explorations of driving culture [5]; it should be evaluated holistically according to what it makes possible and how it does so [6].

II. Related Work

Over two decades ago, Oskar Juhlin presciently 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,

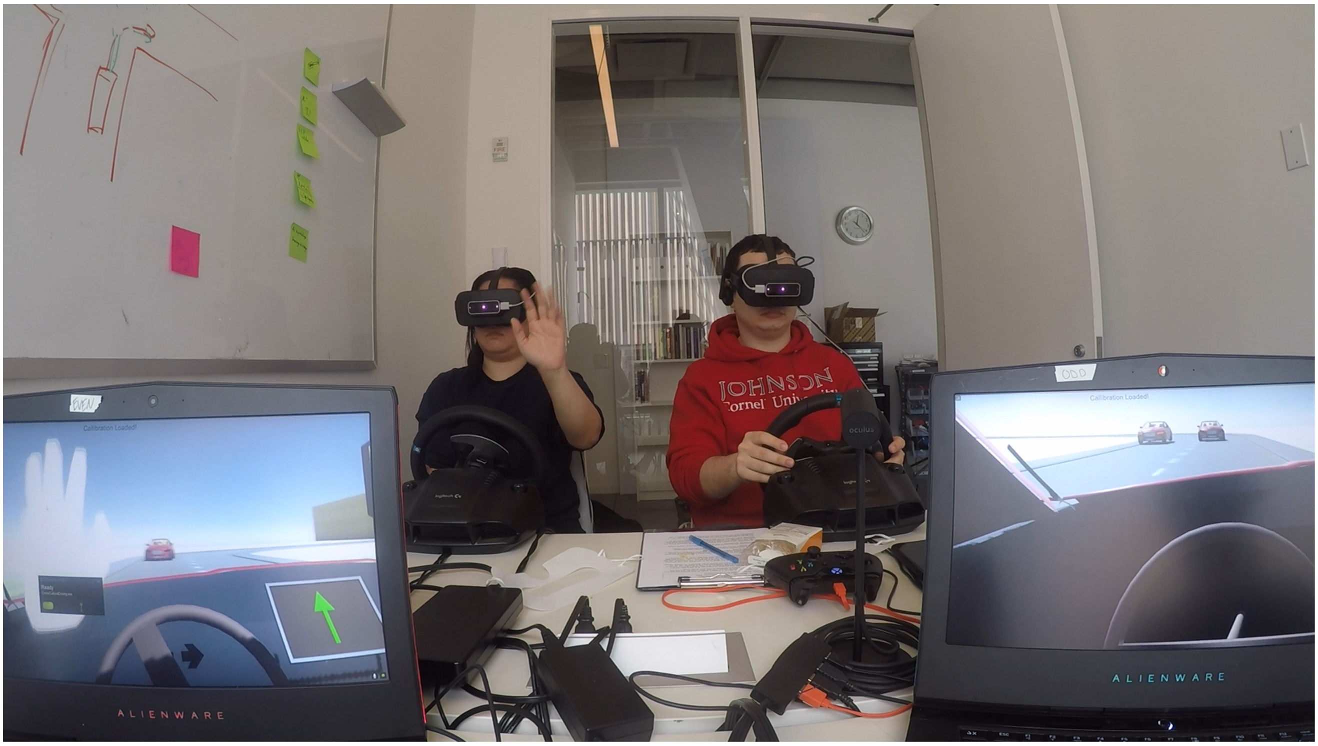


Fig. 1. Our system uses a multi-person virtual reality driving simulation environment to help illuminate how drivers interact in different cultures. The participants are wearing VR headsets with leap motion (hand-tracking device) mounted on the front. Their hands are on the steering wheels and gesturing to each other. Across from them are laptops facing the researchers.

this could lead to ambiguity and misunderstandings, which put serious strains on other road users." [7]

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.

A. Measuring Driving Behavior

While differences in driving across cultures are widely noted, it has proven to be a challenging thing to study; previous research in this domain has focused on profiling driving styles, on self-reports from individuals, and on single-person driving simulation studies targeted at identifying differences in individuals from different cultures.

1) Driving Styles: Sagberg *et al.* define driving styles as habitual ways of driving, which differ across individuals or between groups of individuals. Early research on driving styles aggregated the individual differences of taxi drivers who had been involved in accidents from those who had not, identifying driving style differences that might account for the different accident rates [8].

Specific driving style measures identified by Sagberg include *longitudinal control*, measured by speed, acceleration, jerk, headway distance and time, *lateral control*, measured by lane choice, steering angle, lateral position and lateral acceleration, *gap acceptance*, the time between vehicles at crossings, or passing gap when overtaking, *visual behavior*, the area of fixation, direction of looking, fixation length and frequency, and mirror checking, *errors and violations*, use of indicator, number of infractions, and *other* unusual maneuvers, near accidents, inappropriate honking, gestures made to other users, and driving posture [8].

2) 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 [9]. Using a driver behavior questionnaire [10], 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. "[11]. Driver aggressiveness scales [12] have been used to document differences between Serbian and Romanian drivers [13], driving anger in Spain [14], causes of driving differences of drivers in China [15], and differences in driving between urban and rural U.S. drivers [16]. Driving skill has also been posited to cause the difference between cultures. However, research (also by Özkan *et al.* [17]) examining that hypothesis using the Driver Skill Inventory [18] 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. (Also, 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 [19].)

3) Simulation Studies: Early work in 1989, looking at cross-cultural differences, found that West German simulation drivers were less likely to take risks in a pedestrian crossing scenario than U.S. or Spanish drivers encountering the same situation [20]. These experiments used early simulation environments, where drivers were presented with an overhead schematic of an intersection–like the computer game Frogger–rather than a first-person perspective. Since the same study could be presented to three different populations in three different locations, it could be argued that the differences reveal different tolerances for risk between cultures.

More recently, in 2010, Son *et al.* performed a cross-cultural driving simulation study on older and younger drivers in the US and Korea using a fixed-based immersive driving simulator [21]. Looking at forwarding velocity, speed control, and lateral control measures, this experiment found that US drivers drove more slowly than their Korean counterparts, had a higher range of speed variation, and exhibited better control over lane deviation.

B. *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, right of way rules, but also a performance of attention or inattention [22]. 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 [23].

1) Ethnographic Study: Until recently, most of the research on driving interaction was based on direct or recorded observation. Juhlin, for example, employed ethnographic techniques observing students at a Swedish driving school, interviewing participants, recording driving sessions, and transcribing and thematically coding incidents of cooperation between road users [7]. Similar investigations have been made of social agent navigation in urban traffic [24], driver-bus interaction, [25], pedestrian-vehicle interaction [26], [27] and interactions at petrol stations [28]. Vinkhuyzen and Cefkin used ethnographic techniques to understand how autonomous vehicles will engage with pedestrians, bicyclists, and other cars in a socially acceptable manner and noted the difficulty of making observational distinctions with these methods [29].

2) 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 useful methods for beginning the research, actual simulator studies of driver interaction are recent phenomena [30].

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 [31]. Driving simulators allow experimental control of conditions, reproducibility, ease of data collection, and the ability to test situations that are maybe dangerous to test in real life [32]. 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 largely been made possible through multi-driver simulation platforms. The use of multi-driver simulation studies to examine the interaction between drivers was first performed by Hancock and De Ridder in 2003 [33]. 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.*, in 2011, developed a platform to study interactions between four drivers in a platooning scenario [34]. 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 [35], [36]. 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 [37]. They have published studies using this set up to study cooperative lane-change maneuvers [35], and traffic-light assistance [38] 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 [39].

Other research, particularly targeting other road users, have also used multi-participant simulation setups. A recent publication by Abdelawad *et al.* [43], for example, compares the aforementioned MoSAIC system, the Tokyo Virtual Living Lab networked driving simulation (which is built on OpenStreetMap and CityEngine tools), [41] 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 for hybrid traffic scenario enactments. A recent collaboration between University of Wisconsin-Madison and University of Iowa researchers tested the feasibility of conducting driver-pedestrian simulator experiments with multiple people. [44] To our knowledge, this project is the first proof-of-concept, and we

TABLE I
MULTI-DRIVER SIMULATOR RESEARCH

Multi-driver Simulator Research	Study focus
Hancock and De Ridder 2003 [40]	Collision Avoidance
Muhlbacher 2011 [34]	Platooning
MoSAIC/DLR Heesen et al. 2012[35] Friedrich et al. 2013[36] Oeltze et al.2015 [37] Rittger et al. 2015 [38]	Cooperative driving Overtaking Platooning Traffic light assistance
Tokyo Living Lab Gajananan et al. 2013[41]	Rubbernecking
Houtenbros et al. 2017 [39]	Audio-visual support for intersections
Feierle et al. 2020 [42]	Driver-AV interaction

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 making sure your vehicle does not run into the vehicle in front, the broad field of view is not necessary 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 [45], [46]. In any case, mods such as [46] do not record critical data about the position and behaviors of each driver's car for the purpose of post-analysis and study and have not been validated to produce differences in driving where we expect to see them in regional driving culture.

3) 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 [47], [48], the use of virtual driving simulation for the experimental study of driver behavior is still relatively new [49], [50]. 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 [51].

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 U.K.. They used VR and participant tracking to have two people with VR headsets walk freely across a space play to a game of "Sequential Chicken" with their vehicle avatars in a driving simulation environment [52]. That system illustrates the feasibility of the proposed system in this project, but does not map in-grained driving interaction behaviors to virtual driving as our proposed research would.

III. ARTIFACT DESIGN

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 a portable equipment setup 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.

A. Equipment

The hardware setup for the StrangeLand simulator uses current 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. Conceivably, 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 $2500USD. 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.

B. 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 to synchronize the two clients [53]. This enables participants to see other drivers in their virtual vehicles and note their head orientation or exchange hand gestures.
- *Vehicle model:* We used a prefabricated vehicle model and logic from the GENIVI [54] 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 [55] connecting to SteamVR [56].
- *Hand tracking:* The Orion software pack from Leap Motion for the hand tracking in combination with VR. [57]
- *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.

C. 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 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.

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 [58] (although with slightly different standards about height and placement [59]), 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 clear on the right of way with respect to 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. The resulting set is listed below and shown in Fig. 2:

- **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 four-way intersection scenario.
- **S:3** - *Intersection with Pedestrian:* Both 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 [60].
- **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: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.
- **S:8** is the counterbalanced overtaking scenario.
- **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.

D. 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

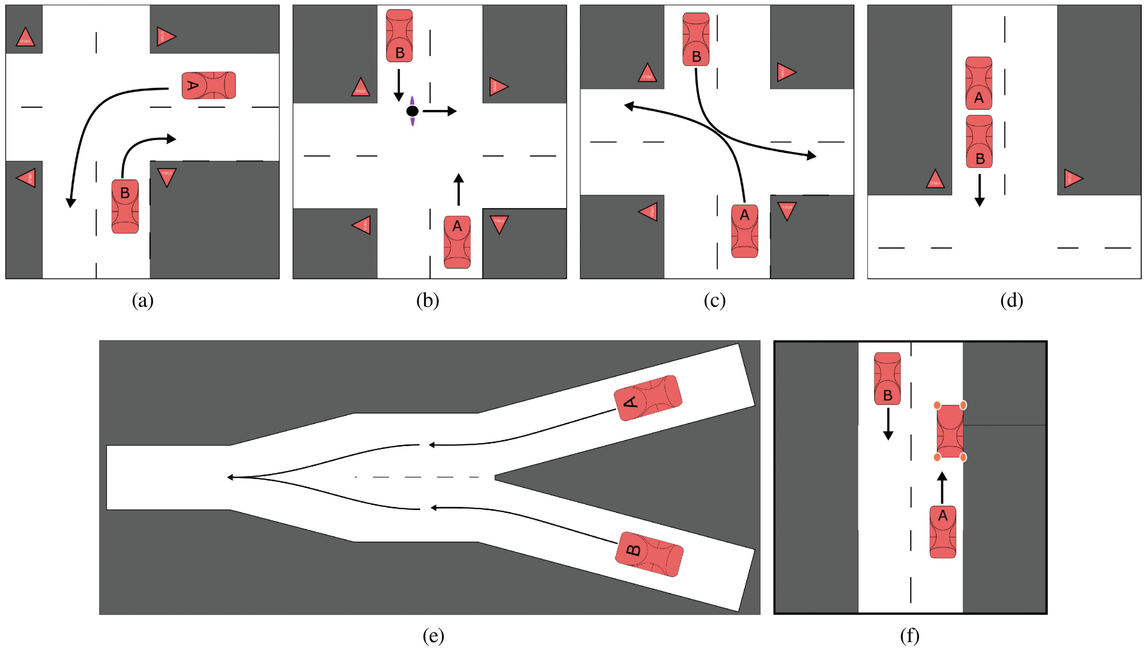


Fig. 2. Showing all the types of scenarios developed for the study. Scenarios that are the exact mirror of another scenario are not shown. Instead, the mirrored version is referenced in parentheses in the image caption. (a) S:1 (S:2) Orth. A Right. (b) S:3 Opposite Pedestrian. (c) S:4 Opposite Left. (d) S:7 (S:8) A Behind B. (e) S:5 (S:6) Merging A Right. (f) S:9 (S:10) A Blocked Lane.

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 (Fig. 3). 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 Fig. 4. In this figure, 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.
- *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 response. E.g., it is easily visible when and how strong 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 [61]. We used this method to avoid taking participants out of the virtual world many times throughout the study. An example is shown in the Fig. 5. 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 the appendix.)

E. 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

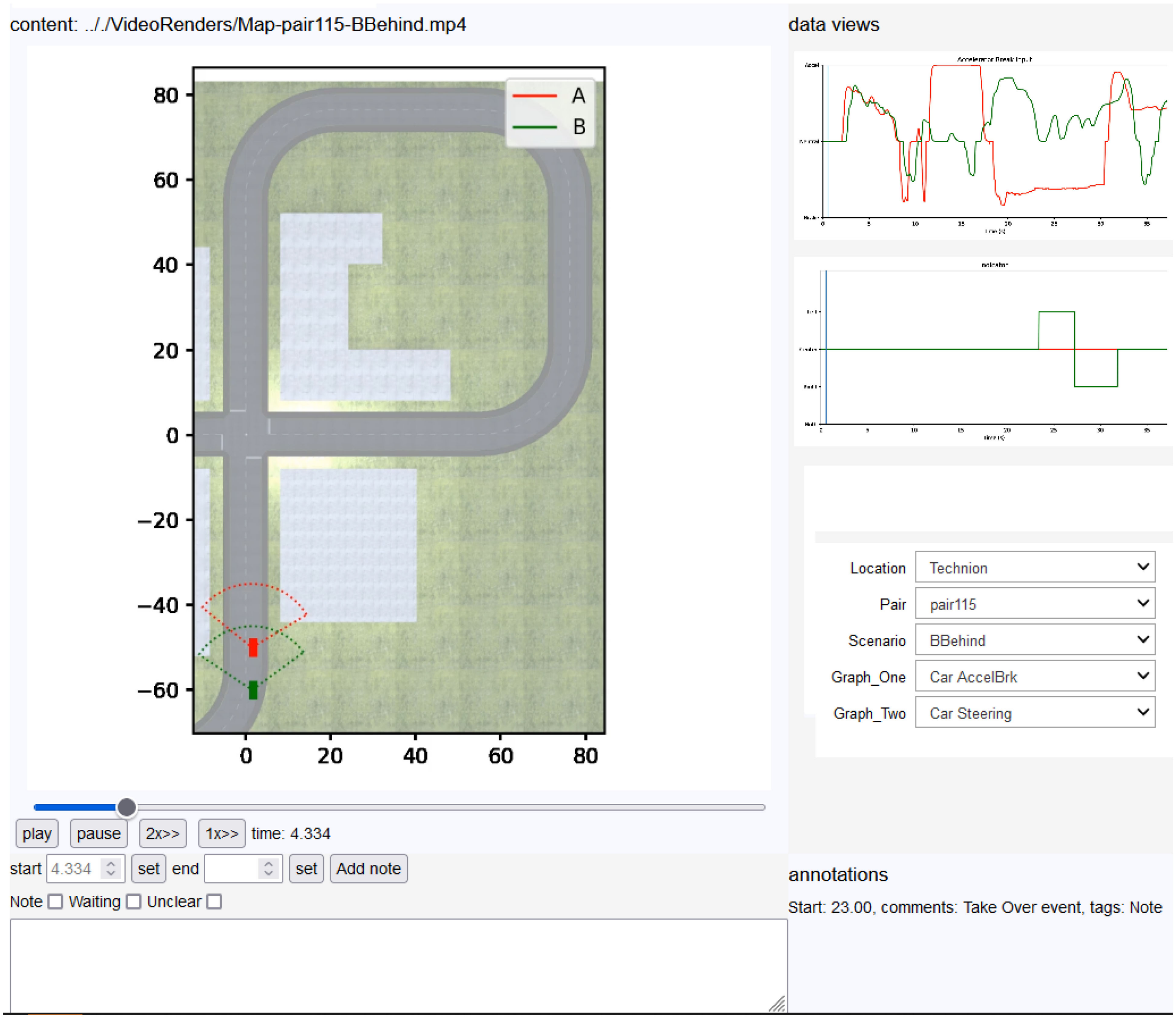


Fig. 3. Screenshot of the behavioral analysis tool. At the top left and top right, we see the map and graphs videos showing the behaviors as they took place. There are several drop-down menus on the right to select what the tool should show (i.e., location of the participants, the participant pair, scenario, and graphs). At the bottom, there is both a space for writing annotations (bottom left) and reading any annotation (bottom right) that has been made for this particular participant pair and scenario.

J2944 Operational definitions of driving performance measures and statistics [62]. 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:* Through the Leap Motion, we collect the hand pose and articulation over time. We can tell if the participants have their hands on the steering wheel, whether they are steering, or if they are waving to someone.
- *Head orientation:* Through the Oculus Rift, 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. [62], [63] 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 orientation:* Through the steering wheel, we collect the rotation information of the steering wheel's position. This, in combination with event logs from the simulation environment, allow us to measure steering reaction time, movement time, and response time, as well as steering reversals.
- *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.

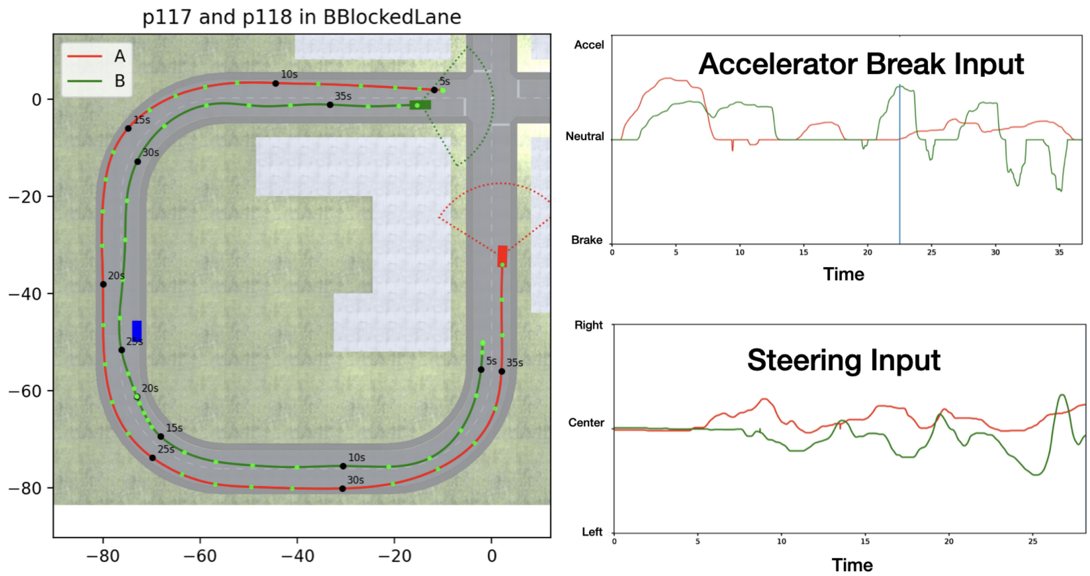


Fig. 4. Top-down view of the virtual environment. This shows the path of two participants in the scenario, Blocked lane. The green tick marks indicate the position over time. The y and x-axes are measured in meters with the intersection at 0/0.

TABLE II
A TABLE SHOWING THE QUESTIONS AND ANSWERS FOR THE CONFIDENCE QUESTIONNAIRE

Confidence Questionnaire	Answer Options					
I felt the other driver drove well.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Does not Apply
I clearly understood the intent of the other driver.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Does not Apply
I felt confident about my own actions.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	

- *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 [62], [64].

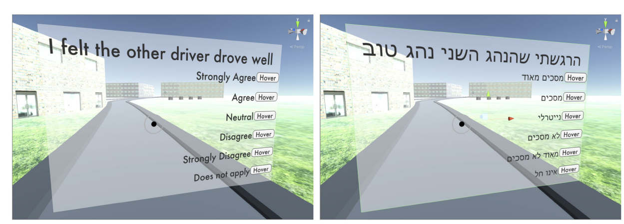


Fig. 5. Post-interaction questionnaire internationalized in (a) English and (b) Hebrew.

IV. 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. For this paper, whose 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 claims about differences in cultural driving behaviors will be attempted and detailed in future work.

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

A. Method

1) 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. Out of the 26 participants, 3 got motion sick 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, if at all, 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, may be 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.

2) Procedure: When participants arrive, 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. [3]

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 Oculus system. 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 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.

B. 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.

1) Deployability: Because our intent is for cross-cultural simulation studies to occur in various locations, the StrangeLand simulator needs to 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 during development, 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 driving interface and the VR headsets–can also be procured for each study site for $< \$1000USD$.

2) 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 our study locations. This was desirable in the proof-of-concept study because we wanted specifically to verify that we were able to 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 is prevalent in unsignalized 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

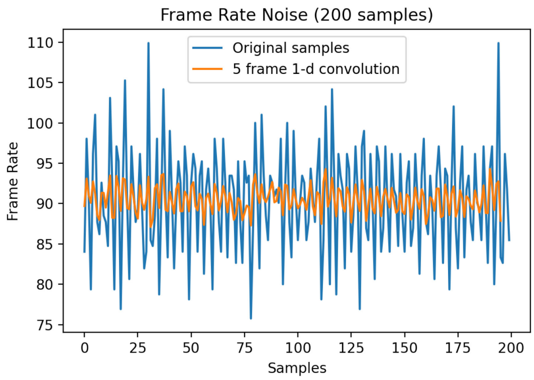


Fig. 6. Showing how the fp/S vary significantly between each frame and how a small 5-frame 1-d convolution shows the continues fp/S.

Fig. 7. We used video recordings with the participants' screen captures in view to validate both the motion-to-photon delay and the network synchronization delay.

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 exist using 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. 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 to skin the cars, buildings, or signs to StrangeLand as they deem appropriate for their own studies. By making the StrangeLand system open-source, and also 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.

3) 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 Fig. 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 $\mathrm{fp/S}$ to create an immersive experience [65]. Data analysis from the proof-of-concept studies showed a median frame rate of 90.9 $\mathrm{fp/S}$ with a standard deviation of 6.0 $\mathrm{fp/S}$. 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.

To ensure that participants did not encounter extended periods of stutter, we computed a 1 d convolution over a window of 4 frames, which reduced the standard deviation by about half to 3.1 $\mathrm{fp/S}$. This finding shows that often a slower frame was preceded with faster frames and that it was exceedingly unlikely($<0.001\%$) for any participant to experience a frame rate of less than 80 $\mathrm{fp/S}$ for more than 5 frames in a row, an excerpt of 200 data points is in Fig. 6.

To both verify motion-to-photon delay and network delay, we used a GoPro action camera (set at a 29.97 $\mathrm{fp/S}$ setting) to take a video recording with both participants in the study setting and their respective virtual views on a laptop screen (see Fig. 7). 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 29.97 $\mathrm{fp/S}$; 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 [66], [67].

4) 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,

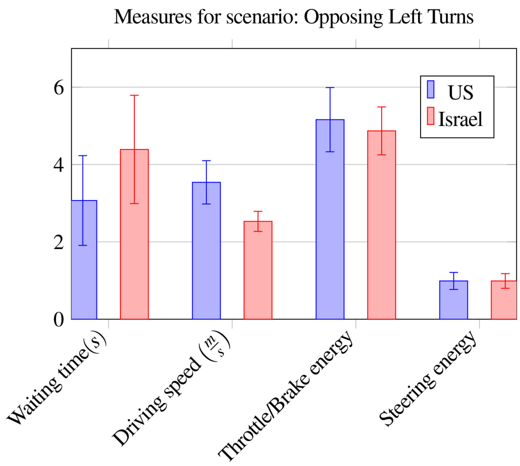


Fig. 8. The comparison of waiting time, driving speed, throttle/brake energy, and steering energy measures for one example scenario illustrate that the system generates consistent data. Error bars show 95% confidence intervals.

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 redesigned 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.

a) Waiting on the other car: We use car position as the reference to all other measures. The Fig. 4 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.

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 Fig. 8 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 participants from Haifa, Israel, and New York, New York. Israelis appear to have driven slower in that specific scenario.

c) 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 Fig. 8 for the Intersection 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 Israel and the US drivers, with Israeli drivers waiting longer.

d) Erraticness of driving: Many of the metrics described in SAE J2944 could be computed based on the available data. However, for the example data, we calculated a simple cumulative difference for the input parameters, longitudinal (speed) and lateral (steering) control. This basic "energy" measurement corresponds to the change in steering and paddle input. This data is exemplary visualized in the last two pairs of bar graphs, in Fig. 8 indicating that Israel and the US participants were comparable in their erratic behavior.

5) 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.

6) 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.

V. Discussion

Our long-term goal in creating the StrangeLand system is to be able 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 work builds on prior work in multi-driver simulation systems, such as [33], [34] 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.

1) Scenarios: The design of interactive systems which can respond to culturally-specific driving interactions requires that an autonomous system recognize interactive bids and maneuvers by people and responds appropriately. This approach–of using simulation environments to elicit naturalistic interactive behaviors–can also help develop other interactive products–such as 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.* have done using data from MIT's Advanced Vehicle Technology dataset [68], [69].

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, the scenarios we developed are not exhaustive; however, when we compare our scenarios with 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 [70].

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 make-up.

2) Cultural Driving Styles: While we do not intend for our proof-of-concept results to be used for broad claims about cultural differences, we believe the findings suggest "construct validity" [71] for cultural driving differences. Our proof-of-concept test found statistically significant cross-cultural differences in driving between the U.S. and Israel drivers in their average recall of turn signals. Additionally, we found a significant difference in average speed between the two countries. 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.

3) Methodological Issues: In running our proof-of-concept test, 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.

4) Features for Studying Interaction: Strange Land 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 a 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 intersection 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.

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

5) Limitations: One key limitation of this work is that the studies in both locations were run with 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 further raises a question on how to correctly screen for a participant from a specific driving culture. i.e., When has someone driven long enough in a specific location to qualify for the experiment.

Over time, when it becomes clearer exactly which measures the 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 [62], 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 [72] 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 [73]–[75]. However, as Mullen *et al.* point out, while simulator driving behavior approximates on-road behavior, it does not replicate on-road behavior [76]. 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.

VI. Conclusion

The differences that we encounter on the road when we travel to new places are no longer fodder for funny anecdotes. The advent of increasing automation makes it important for us to understand the differences we may currently intuit at a deeper, more foundational level. By creating the StrangeLand system, we intend to contribute an experimental platform that allows people from different cultures to experience the same situations and scenarios. This contributes a critical step towards running cross-cultural interaction studies in a controlled and deployable fashion.

Looking ahead, our primary goal is to understand the social interactions and accidents that occur between drivers. With StrangeLand, we look forward to examining interactions between drivers to see how those patterns of interactions are similar or vary across locations. Our system and protocol allow us to compare how people interact on the road and, in so doing, is an essential step towards unlocking the implicit language that drivers are using to communicate with one another on the road.

Acknowledgment

The authors would like to thank Dr. Mishel Johns for his considerable help and advice and Dr. David Sirkin and Dr. Sri Sibi for their support of this project. Thank you to Irene Wei and Xiaoyun Su for helping us run the studies at Cornell Tech. Thank you to Hilla Schefler for her analysis and thank you to (Wendy) Xiaoning Wang for analyzing the head orientation data. Lastly, thank you to Qian Yang for her contributions to the paper.

The simulator with all its assets can be found on Github https://github.com/FAR-Lab/CrossCulturalDriving.

References

[1] X. Harding, "Pitt stop: Inside uber's driverless car experiment," *Popular Sci.*, Sep. 2016. [Online]. Available: https://www.popsci.com/uber-raffi-krikorian-driverless-car/

[2] B. Schoettle and M. Sivak, "A preliminary analysis of real-world crashes involving self-driving vehicles," Univ. Michigan Transp. Res. Inst., Ann Arbor, MI, USA, Rep. UMTRI-2015-34, 2015.

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

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

[5] J. O. Wobbrock and J. A. Kientz, "Research contributions in human-computer interaction," *Interactions*, vol. 23, no. 3, pp. 38–44, 2016.

[6] D. R. Olsen Jr, "Evaluating user interface systems research," in *Proc. 20th Annu. ACM Symp. User Interface Softw. Technol.*, 2007, pp. 251–258.

[7] O. Juhlin, "Traffic behaviour as social interaction-implications for the design of artificial drivers," in *Proc. 6th World Congr. Intell. Transport Syst.*, 1999.
[8] F. Sagberg, G. F. B. Selpi Piccinini, and J. Engström, "A review of research on driving styles and road safety," *Hum. Factors*, vol. 57, no. 7, pp. 1248–1275, 2015.
[9] T. Özkan, T. Lajunen, J. E. Chliaoutakis, D. Parker, and H. Summala, "Cross-cultural differences in driving behaviours: A comparison of six countries," *Transp. Res. Part F: Traffic Psychol. Behav.*, vol. 9, no. 3, pp. 227–242, 2006.
[10] J. Reason, A. Manstead, S. Stradling, J. Baxter, and K. Campbell, "Errors and violations on the roads: A real distinction?," *Ergonomics*, vol. 33, no. 10-11, pp. 1315–1332, 1990.
[11] T. Lajunen, D. Parker, and S. G. Stradling, "Dimensions of driver anger, aggressive and highway code violations and their mediation by safety orientation in UK drivers," *Transp. Res. Part F: Traffic Psychol. Behav.*, vol. 1, no. 2, pp. 107–121, 1998.
[12] J. L. Deffenbacher, R. S. Lynch, E. R. Oetting, and R. C. Swaim, "The driving anger expression inventory: A measure of how people express their anger on the road," *Behav. Res. Ther.*, vol. 40, no. 6, pp. 717–737, 2002.
[13] P. Sârbescu, P. Stanojevic, and D. Jovanovic, "A cross-cultural analysis of aggressive driving: Evidence from serbia and romania," *Transp. Res. Part F: Traffic Psychol. Behav.*, vol. 24, pp. 210–217, 2014.
[14] M. J. Sullman, M. E. Gras, M. Cunill, M. Planes, and S. Font-Mayolas, "Driving anger in Spain," *Pers. Individual Differences*, vol. 42, no. 4, pp. 701–713, 2007.
[15] Y. Ge, W. Qu, C. Jiang, F. Du, X. Sun, and K. Zhang, "The effect of stress and personality on dangerous driving behavior among chinese drivers," *Accident Anal. Prevention*, vol. 73, pp. 34–40, 2014.
[16] J. L. Deffenbacher, "Anger, aggression, and risky behavior on the road: A preliminary study of urban and rural differences 1," *J. Appl. Social Psychol.*, vol. 38, no. 1, pp. 22–36, 2008.
[17] T. Özkan, T. Lajunen, J. E. Chliaoutakis, D. Parker, and H. Summala, "Cross-cultural differences in driving skills: A comparison of six countries," *Accident Anal. Prevention*, vol. 38, no. 5, pp. 1011–1018, 2006.
[18] T. Lajunen and H. Summala, "Driving experience, personality, and skill and safety-motive dimensions in drivers' self-assessments," *Pers. Individual Differences*, vol. 19, no. 3, pp. 307–318, 1995.
[19] S. Amado, E. Arıkan, G. Kaça, M. Koyuncu, and B. N. Turkan, "How accurately do drivers evaluate their own driving behavior? An on-road observational study," *Accident Anal. Prevention*, vol. 63, pp. 65–73, 2014.
[20] M. Sivak, J. Soler, and U. Tränkle, "Cross-cultural differences in driver risk-taking," *Accident Anal. Prevention*, vol. 21, no. 4, pp. 363–369, 1989.
[21] J. Son *et al.*, "Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving," *Int. J. Automot. Technol.*, vol. 11, no. 4, pp. 533–539, 2010.
[22] W. D. Dannefer, "Driving and symbolic interaction 1," *Sociol. Inquiry*, vol. 47, no. 1, pp. 33–38, 1977.
[23] R. Factor, D. Mahalel, and G. 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 Anal. Prevention*, vol. 39, no. 5, pp. 914–921, 2007.
[24] K. Toiskallio, "The impersonal flâneur: Navigation styles of social agents in urban traffic," *Space Culture*, vol. 5, no. 2, pp. 169–184, 2002.
[25] D. Normark, "Enacting mobility. Studies into the nature of road-related social interaction," Ph.D. dissertation, Univ. Gothenburg, Gothenburg, Sweden, 2006.
[26] M. Šucha, "Road users' strategies and communication: Driver-pedestrian interaction," in *Proc. Transport Res. Arena*, 2014.
[27] D. Dey and J. Terken, "Pedestrian interaction with vehicles: Roles of explicit and implicit communication," in *Proc. 9th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2017, pp. 109–113.
[28] D. Normark, "Tending to mobility: Intensities of staying at the petrol station," *Environ. Plan. A*, vol. 38, no. 2, pp. 241–252, 2006.
[29] E. Vinkhuyzen and M. Cefkin, "Developing socially acceptable autonomous vehicles," in *Proc. Conf. Ethnographic Praxis Industry*, 2016, pp. 522–534.
[30] D. M. Zaidel, "A modeling perspective on the culture of driving," *Accident Anal. Prevention*, vol. 24, no. 6, pp. 585–597, 1992.
[31] J. Lee, N. Ward, E. Boer, T. Brown, S. Balk, and O. Ahmad, "Making driving simulators more useful for behavioral research—simulator characteristics comparison and model-based transformation," National Advanced Driving Simulator, Coralville, IA, USA, Tech. Rep. N2013-016, Oct. 2013.
[32] J. De Winter, P. M. van Leeuwen, and R. Happee, "Advantages and disadvantages of driving simulators: A discussion," in *Proc. Measuring Behav.*, 2012, pp. 47–50.
[33] P. Hancock and S. De Ridder, "Behavioural accident avoidance science: Understanding response in collision incipient conditions," *Ergonomics*, vol. 46, no. 12, pp. 1111–1135, 2003.
[34] D. Mühlbacher, J. Zimmer, F. Fischer, and H.-P. Krüger, "The multi-driver simulator-a new concept of driving simulation for the analysis of interactions between several drivers," *Hum. Centred Autom.*, vol. 1, pp. 147–158, 2011.
[35] M. Heesen, M. Baumann, J. Kelsch, D. Nause, and M. Friedrich, "Investigation of cooperative driving behaviour during lane change in a multi-driver simulation environment," in *Proc. Hum. Factors Ergonom. Soc. Europe Chapter Conf.*, 2012, pp. 305–318.
[36] M. Friedrich *et al.*, "Validation of the mosaic-driving simulator-investigating the impact of a human driver on cooperative driving behavior in an experimental simulation setup," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2013, pp. 2052–2056.
[37] K. Oeltze and C. Schießl, "Benefits and challenges of multi-driver simulator studies," *IET Intell. Transport Syst.*, vol. 9, no. 6, pp. 618–625, 2015.
[38] L. Rittger, D. Muehlbacher, C. Maag, and A. Kiesel, "Anger and bother experience when driving with a traffic light assistant: A multi-driver simulator study," in *Proc. Hum. Factors Ergonom. Soc. Europe*, 2015, pp. 41–51.
[39] M. D. Houtenbos, J. C. de Winter, A. Hale, P. Wieringa, and M. Hagenzieker, "Concurrent audio-visual feedback for supporting drivers at intersections: A study using two linked driving simulators," *Appl. Ergonom.*, vol. 60, pp. 30–42, 2017.
[40] P. A. Hancock, M. Lesch, and L. Simmons, "The distraction effects of phone use during a crucial driving maneuver," *Accident Anal. Prevention*, vol. 35, no. 4, pp. 501–514, 2003.
[41] K. Gajananan, A. Nantes, M. Miska, A. Nakasone, and H. Prendinger, "An experimental space for conducting controlled driving behavior studies based on a multiuser networked 3D virtual environment and the scenario markup language," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 4, pp. 345–358, Jul. 2013.
[42] A. Feierle, M. Rettenmaier, F. Zeitlmeir, and K. Bengler, "Multi-vehicle simulation in urban automated driving: Technical implementation and added benefit," *Information*, vol. 11, no. 5, 2020, Art. no. 272.
[43] K. Abdelgawad, J. Gausemeier, R. Dumitrescu, M. Grafe, J. Stöcklein, and J. Berssenbrügge, "Networked driving simulation: Applications, state of the art, and design considerations," *Designs*, vol. 1, no. 1, p. 4, 2017, doi: 10.3390/designs1010004.
[44] D. A. Noyce and J. K. Kearney, "Multi-Modal Distributed Simulation Combining Cars, Bicyclists, and Pedestrians," pp. 1–22, Sep. 2018. [Online]. Available: https://www.researchgate.net/publication/338716277_Multi-Modal_Distributed_Simulation_Combining_Cars_Bicyclists_and_Pedestrians
[45] J. Tilley, "Grand theft auto V: The rise and fall of the diy self-driving car lab," *Forbes*, Oct. 2017. [Online]. Available: 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
[46] "Gta v + universe [archived page]," Jan. 2017. [Online]. Available: https://web.archive.org/web/20170112173923/https://openai.com/blog/GTA-Vplus-Universe/
[47] S. Bayarri, M. Fernandez, and M. Perez, "Virtual reality for driving simulation," *Commun. ACM*, vol. 39, no. 5, pp. 72–76, 1996.
[48] A. Kemeny, "From driving simulation to virtual reality," in *Proc. Virtual Reality Int. Conf.*, 2014, pp. 1–5.
[49] S. M. Taheri *et al.*, "Virtual reality driving simulation for measuring driver behavior and characteristics," *J. Transp. Technol.*, vol. 7, no. 2, pp. 123–132, 2017.
[50] Q. C. Ihemedu-Steinke, D. Sirim, R. Erbach, P. Halady, and G. Meixner, "Development and evaluation of a virtual reality driving simulator," *Mensch Computer 2015-Workshopband*, 2015, pp. 491–498.
[51] S. Cao, K. Nandakumar, R. Babu, and B. Thompson, "Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display," *Virtual Reality*, vol. 24, no. 3, pp. 503–513, 2020.
[52] F. Camara *et al.*, "Pedestrian models for autonomous driving part ii: High-level models of human behavior," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 9, pp. 5453–5472, Sep. 2021.
[53] Unity Technologies, "Unity 3D," May 2019. [Online]. Available: https://unity3d.com/unity/whats-new/2018.4.0
[54] GENIVI Alliance, "GENIVI driving simulator," Jun. 17, 2017. [Online]. Available: https://github.com/GENIVI/genivi-vehicle-simulator

[55] Valve Software, "OpenVR," May 2019. [Online]. Available: https://github.com/ValveSoftware/openvr

[56] Valve, "Steam," 2003. [Online]. Available: https://store.steampowered.com/app/250820/SteamVR/

[57] UltraLeap "Leap motion orion SDK," 2016. [Online]. Available: https://developer.leapmotion.com/orion

[58] Wikipedia, "Yield sign," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Yield_sign

[59] Gov.U.K., "The highway code, road safety and vehicle rules," 2009. [Online]. Available: https://www.gov.uk/browse/driving/highwaycode-road-safety

[60] SafeMotorist.com, "Who has the right of way?," 2022. [Online]. Available: http://www.safemotorist.com/Articles/Right_of_Way.aspx

[61] M. R. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proc. IEEE Nat. Aerosp. Electron. Conf.*, 1988, pp. 789–795.

[62] Society of Automotive Engineers, "Operational definitions of driving performance measures and statistics," *Surface Vehicle Recommended Practice J2944_201506*, SAE Warrendale, PA, 2015. [Online]. Available: https://www.sae.org/standards/content/j2944_201506

[63] S. E. Lee *et al.*, "A comprehensive examination of naturalistic lane-changes," Nat. Highway Traffic Saf. Admin., Washington, DC, USA, Tech. Rep. DOT HS 809 702, 2004.

[64] O. Nakayama, T. Futami, T. Nakamura, and E. R. Boer, "Development of a steering entropy method for evaluating driver workload," *SAE Trans.*, vol. 108, pp. 1686–1695, 1999.

[65] P. Fuchs, *Virtual Reality Headsets-A Theoretical and Pragmatic Approach*. Boca Raton, FL, USA: CRC Press, 2017.

[66] L. Pantel and L. C. Wolf, "On the impact of delay on real-time multiplayer games," in *Proc. 12th Int. Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2002, pp. 23–29.

[67] M. Claypool and K. Claypool, "Latency and player actions in online games," *Commun. ACM*, vol. 49, no. 11, pp. 40–45, Nov. 2006.

[68] J. E. Domeyer, J. D. Lee, H. Toyoda, B. Mehler, and B. Reimer, "Interdependence in vehicle-pedestrian encounters and its implications for vehicle automation," *IEEE Trans. Intell. Transp. Syst.*, to be published. doi: 10.1109/TITS.2020.3041562.

[69] L. Fridman *et al.*, "MIT advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation," *IEEE Access*, vol. 7, pp. 102 021–102038, 2019.

[70] G. Markkula *et al.*, "Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic," *Theor. Issues Ergonom. Sci.*, vol. 21, no. 6, pp. 728–752, 2020.

[71] L. J. Cronbach and P. E. Meehl, "Construct validity in psychological tests," *Psychol. Bull.*, vol. 52, no. 4, pp. 281–302, 1955.

[72] G. L. Wells and P. D. Windschitl, "Stimulus sampling and social psychological experimentation," *Pers. Social Psychol. Bull.*, vol. 25, no. 9, pp. 1115–1125, 1999.

[73] H. C. Lee, D. Cameron, and A. H. Lee, "Assessing the driving performance of older adult drivers: On-road versus simulated driving," *Accident Anal. Prevention*, vol. 35, no. 5, pp. 797–803, 2003.

[74] O. Shechtman, S. Classen, K. Awadzi, and W. Mann, "Comparison of driving errors between on-the-road and simulated driving assessment: A validation study," *Traffic Inj. Prevention*, vol. 10, no. 4, pp. 379–385, 2009.

[75] D. R. Mayhew, H. M. Simpson, K. M. Wood, L. Lonero, K. M. Clinton, and A. G. Johnson, "On-road and simulated driving: Concurrent and discriminant validation," *J. Saf. Res.*, vol. 42, no. 4, pp. 267–275, 2011.

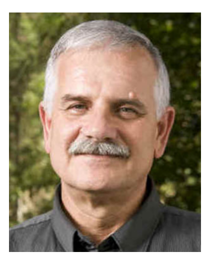
[76] N. Mullen, J. Charlton, A. Devlin, and M. Bedard, "Simulator validity: Behaviours observed on the simulator and on the road," in *Handbook of Driving Simulation for Engineering, Medicine and Psychology*. Boca Raton, FL, USA: CRC Press, 2011, pp. 1–18.

David Goedicke (Graduate Student Member, IEEE) received the M.Sc. degree in human media interaction from the University of Twente, Enschede, The Netherlands, in 2017. In 2018, he joined Cornell's Information Science Ph.D. Program, where he continued his work on developing purpose build human subject simulators to explore how people and automation might interact in the future. His research focuses on exploring how contextual sounds could be used to more intently design machine actions for smoother interaction with people.

Carmel Zolkov received the bachelor's degree in industrial engineering and the M.Sc. in industrial engineering from the Technion, Haifa, Israel. Her academic background is in the fields of industrial engineering, cognitive science, and HCI. She is currently a Data Science Product Analyst. Her research focuses on symbols in support of depth and distance estimation and its implications to augmented reality display.

Natalie Friedman is currently working toward the Ph.D. degree with Cornell Tech, New York, NY, USA. She has a background in cognitive science and HCI. She currently researches how humans interact with robots. Her research focuses on improving human-robot interaction through designing clothing.

Talia Wise is currently working toward the M.Sc. adegree with Technion's Industrial Engineering and Management Faculty, Israel. She has a background in computer science and liberal arts. Her research focuses on how to spark creative idea generation through algorithmically generated word recommendations.



Avi Parush is currently an Associate Professor with Industrial Engineering and Management Faculty, The Technion, Israel, an Adjunct Professor with Queen's University, Kingston, ON, Canada, the University of Victoria, BC, Canada, and an Emeritus Professor from Carleton University, Ottawa, ON, Canada. With an academic background in cognitive experimental psychology, Avi's areas of expertise are human factors engineering, human computer interaction, and usability engineering. His professional and academic career of close to 40 years in human factors is devoted to influencing the design of workplaces and tools people use in order to make their lives easier, safer and more empowered. His current research interests include teamwork in complex and critical situations, human factors in healthcare, human—intelligent systems integration, and human situational awareness in various contexts. He is the emeritus founding Editor in Chief of the *Journal of Usability Studies,* an Associate Editor for the *Human Factors in Healthcare journal*, and on the Editorial Board of the *Human Factors Journal*. Avi moutain bikes and strives to leave a better world to his children and grandchildren.

Wendy Ju received the master's degree in media arts and sciences from MIT, and the Ph.D. degree in mechanical engineering from Stanford University, Stanford, CA, USA. She is currently an Associate Professor with the Jacobs Technion-Cornell Institute, Cornell Tech, New York, NY, USA, and the Technion, Haifa, Israel, and with the Information Science field, Cornell University, Ithaca, NY, USA. Her work in the areas of human-robot interaction and automated vehicle interfaces highlights the ways that interactive devices can be designed to be safer, more predictable, and more socially appropriate. Prof. Ju has innovated numerous methods for early-stage prototyping of automated systems to understand how people will respond to systems before the systems are built. .