Evaluating a VR System for Collecting Safety-Critical Vehicle-Pedestrian Interactions

Erica Weng*‡, Kenta Mukoya†, Deva Ramanan*, and Kris Kitani*

*Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15217
†Denso Japan. Work done while visiting at the Carnegie Mellon University Robotics Institute.
‡Correspondence to: eweng@cmu.edu

Abstract—**Autonomous vehicles (AVs) require comprehensive and reliable pedestrian trajectory data to ensure safe operation. However, obtaining data of safety-critical scenarios such as jaywalking and near-collisions, or uncommon agents such as children, disabled pedestrians, and vulnerable road users poses logistical and ethical challenges. This paper evaluates a Virtual Reality (VR) system designed to collect pedestrian trajectory and body pose data in a controlled, low-risk environment. We substantiate the usefulness of such a system through semi-structured interviews with professionals in the AV field, and validate the effectiveness of the system through two empirical studies: a first-person user evaluation involving 62 participants, and a third-person evaluative survey involving 290 respondents. Our findings demonstrate that the VR-based data collection system elicits realistic responses for capturing pedestrian data in safety-critical or uncommon vehicle-pedestrian interaction scenarios.**

Fig. 1: Left: Participant's avatar within the virtual environment from a third-person perspective. Right: participant wearing the VR headset in the data collection environment.

I. INTRODUCTION

Pedestrian trajectory data collection is necessary for training safe and robust pedestrian forecasting models for autonomous vehicles (AVs), but many important vehicle-pedestrian interaction scenarios are missing in existing popular trajectory and tracking datasets [11, 22, 26, 12, 70]. Specifically, collecting data in many scenarios is difficult. For example, it is difficult to collect data of jaywalkers: pedestrians who cross a street at a location other than designated crosswalks or intersections, often in disregard of traffic regulations. Jaywalking is generally considered unsafe and is illegal in many regions, and thus it is sparsely observed in real life or real datasets. However, it is still important that forecasting models understand the behavior of jaywalkers for pedestrian safety. Furthermore,

it is problematic to willfully direct pedestrians to perform illegal behavior for the purpose of data collection. As another example, real-life datasets have a deficiency of data of children pedestrians [11, 22, 26, 12, 70]. Child street-crossing behavior differs from that of adults; children often take more risks and appear more unpredictable [27]. Due to child protection laws, it is even more problematic to put children in potentially dangerous situations for the purpose of data collection [52]. As a final example, there exist very specific vehicle-pedestrian interaction scenarios, such as walking on the road alongside traffic, crossing a parking lot or garage entrance with vehicles entering and exiting, or crossing a narrow urban alleyway in which the building structure obstruct easy view of oncoming vehicles. These scenarios arise due to specific urban structural and geographic features that may only be sparsely found in real-life urban areas. However, these scenarios are still relevant and important to robust pedestrian safety, and thus it is important that data is collected for these scenarios. There is a large body of work on controllable simulators that can generate large quantities of synthetic data of various scenarios of choice [20, 14, 72], including the aforementioned scenarios that are uncommon in real-life but critical to pedestrian safety. There is also much work done on AV perception, prediction, and planning models that utilizes simulated data for training [36, 33]. Relevant to our area of interest, Huang and Ramanan [32] utilize synthetic pedestrian data to train a detector to detect unusual or uncommon pedestrian behaviors. However, all such methods that train largely on synthetic data suffer from the well-known "sim2real" gap. This gap exists because simulators struggle to reproduce certain aspects of reality: for example, in CARLA [20], vehicles travel along "rails" located in the center of the lane, never deviating from the center as real human drivers do. In addition, CARLA's pedestrian models move mechanically according to mathematical gait functions. Due to the complexity of human behavior and movement characteristics, pedestrians are particularly difficult to simulate.

These limitations point to the need for other solutions to obtain such data. Virtual Reality (VR) offers a solution for collecting pedestrian data in a safe environment free of real automotive dangers; VR can be used to immerse a pedestrian in a virtual scenario analogous to one found in real-life, and

the pedestrian can be asked to cross a virtual road, or traverse the virtual environment, all the while their trajectory data is collected. In addition, data collected from VR is collected from real humans, and thus does not suffer from the same human behavioral sim2real gap as simulated data. However, VR may exhibit limitations to realistic data collection. For example, the VR user experience could be unrealistic, whether it's because of the graphics of the virtual environment, lag or delay in the VR view refresh frame rate, or the weight of the VR headset on their head. These issues may interfere with the pedestrian's sense of immersion, and accordingly, may alter their actions and movements from what they would do on a real street. Evaluating a VR system [3] that was developed concurrently with this work through empirical studies, we show that although users do not experience perfect sense of presence or immersion within the system, data collected from the system is mostly indistinguishable from that collected in real life, providing evidence for our claim that the VR system is effective at collecting a genuine pedestrian response to vehicle motion. Our contributions are as follows:

1) Through semi-structured interviews with 4 professionals in the autonomous vehicle (AV) field, we show that a virtual-reality-based system for collecting pedestrian data in traffic scenarios can address deficiencies in existing vehicle-pedestrian datasets and limitations to existing real-life data-collection methods.

2) Through an empirical quantitative and qualitative questionnaire of 62 individuals testing our system, we demonstrate that such a virtual reality (VR) system can elicit a genuine human response to vehicle motion, and result in pedestrian behaviors that mimic those of pedestrians interacting in real-life vehicle-pedestrian scenarios.

3) Through an empirical survey of 290 individuals each answering 3 forced-choice questions comparing the realism of 2D trajectories collected in the VR system vs. in real-life, we measure the similarity of the VR data to real-life data.

II. BACKGROUND AND RELATED WORK

A. Human Behavior in Virtual Environments

Many works have studied human behavior within virtual environments. Some works study the efficacy and quality of the VR experience via objective measures such as stride length and gait [34] as well as subjective measures such as subjects' ratings on presence questionnaires [21, 6, 63, 64]. Some study specific behaviors of interest like normal walking [49], evacuating a building [4], collision-avoidance [7], proxemics and group behavior [56, 50], etc. Some focus on behaviors in specific professional areas like surgery [44, 41] or education [5]. Many methods evaluate the efficacy of VR systems by comparing human behavior in VR vs. real environments [45]. Finally, other works evaluate how different conditions of the VR system's features affect user sense of presence behavior, such as level of photorealism of the

environment rendering [73], locomotion methods [65, 48, 54], and avatar appearance [59, 53].

B. Pedestrian Behavior Studies

Extensive study has been done to characterize pedestrian behavior in traffic environments and responses to vehicle behaviors [35, 15, 39, 24, 69, 13]. Some works study AV behavior or evaluate improvements to the AV-pedestrian communication interface such as intent signalling methods [35, 15, 39, 24, 39, 25, 2, 19, 17]. Many of the aforementioned works, as well as others, use immersive VR environments to simulate scenarios of interest and study pedestrian's responses [51, 60, 8, 8, 23, 43, 67, 51]. Some works focus on evaluating the efficacy of VR systems for use in studying human behavior or evaluating safety systems or vehicle-pedestrian communication systems by comparing subject's responses in virtual environments to that of real-world environments [47, 18, 60, 31]. Work on VR-based dataset collection and benchmarking tools similar to that studied in this work has also been done, such as [16], but their focus is on testing vehicle-pedestrian communication interfaces rather than on collecting realistic vehicle-pedestrian interaction behavior.

C. 2D Pedestrian Trajectory Data Capture: Real-Life, Human-Informed, and Fully-Synthetic

Previous works have used various methods to capture vehicle-pedestrian interaction datasets. The most popular datasets, naturally, are those collected in real-life environments [55, 40, 58]. PedX [37] and PIE [57] contain pedestrian trajectories collected at intersections, which is the closest data to our use case. However, they too are limited in scope; PedX contains full body pose data but in limited environments such as crosswalks, and PIE contains only 2D bounding box annotations. Popular open-source vehicle datasets also contain pedestrians [12, 70**?**]. However, as verified in the first part of our study in Section III, these datasets lack diverse pedestrian behavior and diverse vehicle-pedestrian interactions. Most vehicle-pedestrian interactions in these datasets are at intersection crosswalks.

Some datasets are completely synthetic, collected autonomously via simulated environments that operate based on rules. While some simulators only produce simplistic point trajectory behavior [9, 38], others are extensively customizable and controllable worlds that can model complex sensor data [20, 14, 72]. Although simulators simplify the process of collecting large-scale data in diverse environments, there is a domain gap between synthetic and real data that is often difficult to measure, and many simulators do not reproduce human behavior well. Because AV models ultimately must work for real pedestrians and vehicles, in this way synthetic data falls short.

Some methods, such as driving simulators [62, 46], invite human input to elicit real human behavior. The Garden of Forking Paths [42] collects annotations of trajectory continuations from human annotators to create a multi-future pedestrian trajectory dataset in simulated environments reconstructed

from real-world scenes. However, they only annotate trajectories at the 2D point-level, through a very limited keyboard and mouse interface, a far cry from the controllability of a VR headset and environment.

III. PART 1: UNDERSTANDING LIMITATIONS OF EXISTING VEHICLE-PEDESTRIAN DATASETS

In order to better substantiate the use cases for a VR pedestrian data-collection system, we sought to confirm the apparent limitations of existing vehicle-pedestrian interaction datasets with AV professionals. Thus, in the first part of our study, we conducted semi-structured interviews with professionals who had experience working with vehicle-pedestrian trajectory datasets. Our goal was to answer the research question: *What are the limitations of existing vehicle-pedestrian datasets with respect to the availability and quality of pedestrian data?*

A. Semi-Structured Interview Recruitment and Design

For the interviews, we recruited 3 academic researchers and 1 industry practitioner through direct contacts. Two academic participants have performed research in the AV perception, prediction, and planning stack (referred to by I1 and I2). The other academic participant performs research in social navigation for mobile robots that interact closely with humans (I3). The industry participant, meanwhile, is a systems test engineer at a large AV company (I4). During the study, all participants self-reported that they had experience working with vehicle-pedestrian interaction data.

In our interviews, we first asked participants to describe their area of expertise; Then, we asked them to describe their understanding of the current state of vehicle and pedestrian trajectory datasets: *Which trajectory datasets you have worked with? What do you feel are the current limitations of these datasets?* We probed deeper into the state of current vehicle-pedestrian trajectory datasets with questions such as *What kinds of scenarios are lacking in real datasets?* and *What existing methods are there for improving performance in uncommon or out-of-distribution scenarios?* Finally, we asked directly for their opinion about the potential limitations and benefits of data collected in a virtual environment via a VR simulator. We asked questions such as *Given a Virtual Reality system in which a pedestrian walks around while wearing a VR headset in a virtual traffic environment while sensors capture their movements: what are the potential benefits and limitations of such a system?*

B. Interview Responses

To analyze the state of the limitations of current trajectory datasets, we used inductive thematic analysis [10], grouping the themes that arose during the interviews into four distinct categories, which we describe below.

1) Lack of Uncommon but Important Scenarios: Vehicle-pedestrian datasets lack scenarios that are important to AV safety, but rare in-the-wild or difficult to collect data for. I2 commented that vehicle-pedestrian datasets contain plenty of data of pedestrians crossing roads at intersections, but little data of pedestrians walking on the sidewalk alongside the road. These scenarios are just as important as pedestrians crossing at crosswalks, because pedestrians walking alongside the road could become jaywalkers in the near future.

2) Lack of Interesting Vehicle-Pedestrian Interactions: Both I1 and I2 pointed out that the most popular public vehicle-pedestrian datasets lack "interesting interactions." One of the most popular datasets [12, 70] in particular contains a subset that was compiled specifically for the "interestingness" of its scenarios based on internally-defined heuristics, such as number of other agents present in the scene, speed changes, and lane changes. However, I1 remarked that these scenarios were still mostly uninteresting, claiming that it is difficult to come up with a straightforward heuristic to separate "interesting" scenes from "boring" scenes. I2 stated that around 75-80% of the data they worked with is "uneventful", and that there was insufficient diversity in scenes. Scenes often featured empty or minimally populated roads where only the ego-vehicle, the vehicle performing the data collection, was present. Vehicles also lacked varied behaviors such as lane changes. I3 claimed that the pedestrian-only datasets they worked with also lacked variety in environment layouts and did not contain many pedestrian interactions.

3) Lack of Fine-grained Trajectory Features: Vehicle-pedestrian datasets lack richness in representation. I1 commented that existing trajectory datasets are heavily preprocessed from raw lidar data, losing a lot of the richness of the raw representation. Popular trajectory datasets such as Argoverse [12, 70] and NuScenes [11] represent pedestrian trajectories as 2D points rather than full 3D human bodies. I2 notes that while pedestrian body orientation is also recorded, it still falls far from the detail available in full body pose. In tracking datasets, agents are often annotated with simple 3D rectangular bounding boxes. I3 explained that pedestrian bounding boxes often overlap with one another, because humans occupy their bounding boxes only sparsely. Overlapping bounding boxes appear to be colliding; this may lead to model training issues such as failure to understand collision-avoidance. Thus, bounding boxes also fall short in representing the richness of pedestrian motion.

Pedestrian body skeleton pose and head direction can inform understanding of pedestrian intent high-level pedestrian motion. For example, pedestrians often look both ways before they cross the street, and a pedestrian that is looking in the opposite direction of an oncoming car is less likely to stop for that car, as they do not see it. In addition, traffic officers often stand in the middle of the road to direct traffic with their arm movements. However, as I4 pointed out, their company's AV system has limitations in predicting the behavior of traffic officers, often mis-predicting them as pedestrians crossing the road rather than static pedestrians who will not intercept the vehicle's path in the near-future.

4) Summary: The shortcomings of current datasets point to the need for additional data solutions to supplement existing datasets. The VR pedestrian simulator we evaluate in this paper can be manipulated to recreate diverse and uncommonly
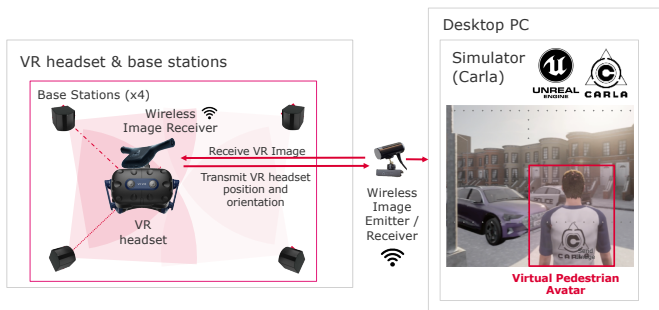
Fig. 2: Schematic of the VR data collection system.

Fig. 3: Left: The VR system set up in the classroom data collection space. Right: A pedestrian subject wearing the VR headset.

seen traffic scenarios, addressing the first two themes. It can also record body pose information, addressing the third theme. There were some concerns brought up by our interview participants that data collected from a VR system may have a domain gap with real-world data, just as other simulators do. In the next two sections, we provide evidence that in spite of the domain gap, the VR environment still elicits a genuine and realistic pedestrian response.

IV. VR PEDESTRIAN TRAJECTORY AND POSE COLLECTION SYSTEM OVERVIEW

We briefly describe the VR pedestrian data-collection system that we evaluated.

The simulator uses the HTC VIVE Pro 2 as its VR system. Steam VR is used to interface the HTC VIVE devices with a controllable game engine, so they can be used as input devices within the simulation. The simulator is built on top of the CARLA simulator using Unreal Engine. Since Steam is only available on Windows machines and the HTC VIVE Wireless Adapter transmitter must be connected to the computer via an RP-SMA port, the Alienware Aurora R13 Gaming Desktop is a suitable machine for use as the underlying server. The system specifications are as follows: 12th Intel core i9 CPU, NVIDIA RTX 3080 GPU, 64 GB memory, M.2 SSD 2TB storage, Windows 11 Home OS.

The data collection environment requires a rectangular collection space free of obstacles measuring approximately 40'x 20' in dimensions. Four HTC VIVE base stations are set up in the four corners of this space. Eight GoPro Hero 10 cameras are set up along the two long edges of the space, which record the subject from different angles so that their skeleton pose can be extracted out later. A schematic of the VR system components is shown in Figure 2, and the actual data collection space with system setup is shown in Figure 3.

Because the engineering and design of the VR pedestrian simulator is not the primary contribution of this paper but of [3], we leave further details about the data collection system to that work.

V. PART 2: VR SENSE OF PRESENCE USER STUDY

To determine the degree to which users felt immersed and present in the virtual world,

A. Study Design

We recruited participants by word-of-mouth, university email lists, and physical flyers posted around the university campus. Participants with physical disabilities, vulnerable subjects, and minors were excluded from participation. Participants came to the university campus to participate in the study, where we set up the data collection environment as described in the previous section in a large, empty classroom.

After signing an informed consent form, participants were asked to put on the VR headset and to familiarize themselves with the VR headset and virtual environment by walking around in a version of the simulation without moving vehicles. Then, we asked participants to complete 3 tasks, each of which featured a different traffic environment in which they had to walk to reach a goal destination while in close interaction with moving vehicles: 1) jaywalking across a two-way street, 2) walking alongside moving vehicles on the road, and 3) crossing a crosswalk at a 4-way intersection with stop signs. The 3 tasks are depicted in Figure 4. To inform participants of their goal destinations, we placed colored square markers on the ground within the virtual environment, and used commands such as "Do you see the colored square on the ground on the other side of the road? Please walk to it" to direct the user. After completing all 3 tasks, participants were asked to complete a presence questionnaire to evaluate their experience.

B. Self-Reported Measures

To design the questions asked in the post-experience questionnaire, we used a combination of questions derived from previous works, questions modified from previous works, and custom questions designed specifically to evoke defining attributes of the VR user experience. We designed 12 quantitative questions using a semantic differential scale [66] from 1 to 7, as well as 4 free-response questions with free text entry. We grouped the questions into 3 categories reflecting different aspects of trustworthiness of the data collection system: sense of presence experienced within the virtual environment [6, 71] (coded by the letter P), sense of agency (A), and behavioral and experiential similarity to real-life (B). In total, the post-condition questionnaire was composed of 16 questions aimed to cover a range of subjective ratings while keeping the time for participants to complete the questionnaire within 10 minutes. Questions, sources, categorizations, and semantic differential anchors for the questions are recorded in Table I.
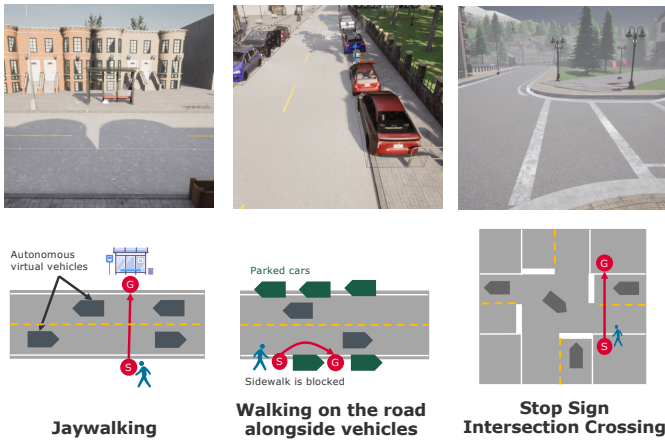
Fig. 4: The three tasks that we ask our study participants to complete. Top: snapshot of each task as seen in the CARLA virtual environment. Bottom: schematic of each task; "S" denotes participant start location and "G" denotes goal. Schematic images also used in Anonymized [3].
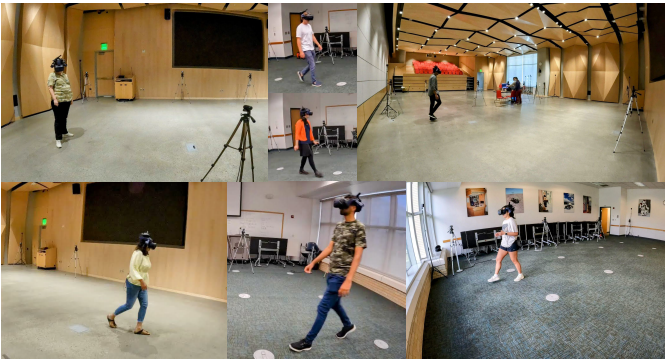

Fig. 5: Various study participants wearing the VR headset and walking around in two different classroom-setting data collection environments.

Participants also reported various demographics: age, gender, VR experience (5-point Likert-type scale), video game frequency (5-point Likert-type scale), frequency of jaywalking behavior (5-point Likert-type scale), and level of alertness (Stanford Sleepiness Scale [61]). Summary of participant demographics is reported in Table II.

C. Results and Analyses

1) Quantitative Measures: A total of $N=62$ participants participated in our study. Aggregating all responses across all questions in a category, users rated their sense of presence (P) 5.6, sense of agency (A) 5.5, and similarity of their behavior and experience to that of the real world (B) 5.3. These numbers suggest that users experience a relatively moderate-high sense of presence, agency, and behavior similarity to the real world. The users' aggregate response of 5.3 in the behavioral similarity category is a little lower than that in the sense of presence category and sense of agency category, possibly suggesting that high sense of presence is not sufficient

to guarantee that a pedestrian behaves exactly the same in VR as they do in the real world. One possible explanation for this is that, for the high proportion of users who rarely used VR, the novelty factor of the VR environment causes them to behave differently, even though they are experiencing a high degree of presence and immersion.

We report the average user ratings for each question in Table III.

2) Qualitative Measures: The qualitative responses evoke greater detail from the participants' experiences. In response to the free-response question, *What aspects of the systems or environment were realistic?*, participant responses fell into the following major categories:

1) Traffic patterns and flow were very authentic.
2) The visual rendering contributed to a realistic experience: the relative sizes and dimensions of cars, buildings, roads, sidewalks, trees and other aspects of the environment felt accurate, appropriately sized, and well-rendered. The environment felt realistically designed. Some users commented that the curb appeared so realistic that it caused them to stumble when they actually tried to step onto it, as no physical curb existed in real-life.
3) Movement of images along the line of vision was smooth when the user moved their head or body.
4) Cars moved and behaved realistically with respect to speed, positioning, and timing.
5) The environment elicited emotions and caution similar to real-world experiences, such as feeling threatened by cars and being conscious of making mistakes, like crossing the road at the wrong time.

In response to the question, *What parts of the system or environment were unrealistic?*, the main concerns brought up by participants include:

1) Dizziness and nausea while using the system
2) System stability issues: lag, flicker, and instances in which the virtual space would re-calibrate and the participant would be transported within the virtual environment to a different location despite not having moved
3) Awareness of the real world and obstacles in the real world, which took away from sense of presence
4) Lack of peripheral vision due to the construction of the VR headset goggles, which resulted in participants turning more to the left and right to check for vehicles than they do in real life
5) Lack of environmental sound
6) Imperfect visual information: lack of stop signs and walk signals at intersections, game-asset quality of buildings and surroundings, not all body parts visible in the simulation
7) Imperfect tactile information: sidewalk curb observed in virtual environment, but real-world environment lacked a tactile height change
8) Mechanical and unnatural movement of vehicles, which did not deviate from fixed paths nor yielded to pedestri-

ID	Source	Category	Question	Semantic Differential Anchors (1→ 7)
P1	SUS [68]	presence	Please rate your sense of "being there" in the environment.	did not have a sense of "being there" → normal experience of "being there"
P2	SUS [68]	presence	When you think back on your experience, do you think of the environment more as images that you saw, or more as somewhere that you visited?	Images that I saw → somewhere that I visited
P3	WS [71]	presence	How real did the objects in the environment seem?	very fake, clearly images → very real, like I could touch them
P4	custom	presence	During the VR experience, were you more concerned with the real world (this classroom) or the virtual world?	real world → virtual world
P5	SUS [68]	presence	When you think back on your experience, do you think of the vehicles more as images that you saw, or as things you interacted with?	Images that I saw → things that I interacted with
P6	custom	presence	During the experience, did you often think to yourself that the vehicles were physical objects that could have actually hit you and caused you injury?	very much so → not at all
A1	custom	agency	How comfortable did you feel moving around in the environment?	very uncomfortable → very comfortable
A2	AE [28]	agency	How much did you feel like you could control the virtual body?	did not feel much agency → could control it like own body
A3	SPES [30]	agency	How freely did you feel you could move in the environment?	movements were restricted → movements were free
B3	NASA-TLX [29]	behavior	How mentally demanding were the tasks compared to doing them in the real world (e.g. on a real street)?	less demanding → more demanding
B1	custom	behavior	Did you feel your head, arm, and body movements were the same as they would have been in the real world?	completely different → exactly identical
BF1	custom	behavior	What parts of your movements were different than how they would have been in the real world?	*free response*
B2	custom	behavior	Did you feel your decisions about when to act were the same as they would have been in the real world?	completely different → exactly identical
BF2	custom	behavior	What parts of your decisions were different than how they would have been in the real world?	*free response*
PF1	custom	presence	What aspects of the systems or environment were realistic?	*free response*
PF2	custom	presence	What aspects of the system or environment were unrealistic?	*free response*

TABLE I: Post-condition questionnaire used in the study.

Variable	N=64
Age {18-37} (years)	24.71 (4.17)
% Male	63% (n=39)
% Female	37% (n=23)
Video gaming frequency {1-7}	3 {1-5}
VR experience {1-7}	2 {1-4}
Jaywalking frequency {1-7}	3 {1-5}
Level of alertness {0-7} [61]	6 {3-7}

TABLE II: Summary of participant demographics. Continuous variables are summarized as *mean* (*standard deviation*) and ordinal variables are summarized as *median* {*range*}.

Question Category	Presence						Agency			Behav. Sim.		
ID (as in Table I)	P1	P2	P3	P4	P5	P6	A1	A2	A3	B1	B2	B3
Rating	5.9	5.3	5.0	6.2	5.9	5.3	5.1	5.6	5.9	5.0	5.3	4.8
Average			5.6					5.5			5.3	

TABLE III: VR Presense Questionnaire Evaluation Results.

ans

9) Lack of drivers in vehicles, which rendered users unable to make eye contact with drivers to determine when to cross
10) Lack of other pedestrians in the environment
11) Being told where and when to start and stop, unlike in real life where this is self-determined

In response to the question, *What parts of your movements were different than how they would have been in the real world?* participant responses group into the following major concerns:

1) Did not turn head as much due to weight of the VR headset and limited peripheral vision
2) step movements differed due to difference between vi-

sual perception and tactile perception (such as the curb)
3) Walking style: keeping hands in front of body as a defensive body posture to avoid bumping into walls in the real world, or to fend off aggressive vehicles in the virtual world
4) Navigation around static objects in the virtual world such as parked cars differed due to the unexpected dynamics of those objects, such as moving aside when users collided with them

In response to the question, *What parts of your decisions were different than how they would have been in the real world?* subjects reported the following major themes:

1) Some subjects were more hesitant to move in the virtual environment than in real life, due to the realism limitations discussed in the responses to the question above.
2) Other subjects had an added sense of urgency to act, because of unexpected or erratic vehicle behavior, or because lack of peripheral vision limited the subjects' awareness of cars
3) Some subjects exhibited more risky behavior due to less fear about being hit by a virtual car.

The qualitative responses from our user study affirm that, while the system does have areas for improvement, there are noteworthy strengths. Participants largely found the visual aspects and their own emotional responses to be authentic to their real-world experience. Some users did report aspects of the virtual environment that seemed unrealistic; yet sense of presence, agency, and behavioral similarity to real-life still seems high (Table III). This seems to suggest that complete realism of the virtual environment may not be necessary to have high sense of presence,

Furthermore, for users that found their behavior or movements differing from that in real life, the different behaviors are

not necessarily less genuine than those of real life. Uncommon, edge-case scenarios sometimes do not feel real because they are unexpected, but they are no less important for ensuring the safety of AVs. The VR system is exactly designed to evoke pedestrian response in those edge-case scenarios, thus improving the collection of data in those scenarios.

VI. PART 3: THIRD-PERSON REALISM EVALUATION OF VR VS. REAL-LIFE VS. SYNTHETIC 2D TRAJECTORY DATA

First-person subjective evaluation provides valuable insights into how well the VR system evokes realistic behavior. However, to further substantiate the comparability of real-life and VR data, we perform an additional third-person evaluation evaluative study. In this study, we asked external evaluators to try to distinguish between trajectory data collected in real-life vs. VR. This study evokes a different facet of trajectory realism, one that comes not from first-hand experience, but from third-person evaluation.

A. Study Design

To collect real-life jaywalking trajectories, we visited a two-way single-lane street near the university that is well-trafficked by jaywalkers as well as vehicles. We chose this setting for its potential to provide real-life trajectories in a similar geographic layout to that we used in the VR jaywalking scenario. In contrast to the VR scenario, there were fewer vehicles as well as more irregular gaps between vehicles, making it easier for jaywalkers to find a gap in which to cross. A DJI Mavic Mini drone [1] was used for data collection, flown at a height of approximately 25 meters. In a 1-hour time interval, we secured footage for 7 jaywalkers.

The trajectories of these jaywalkers were manually annotated via a trajectory annotation tool and interpolated to 20 frames per second to match the frame rate of the trajectory data collected from the VR system. The resulting trajectories were smoothed via Gaussian smoothing with a standard deviation of 0.35 seconds, as the VR system also performs some smoothing to eliminate noise from the collected trajectories.
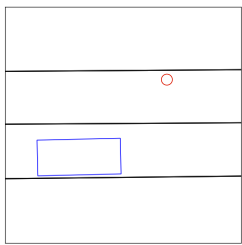

Fig. 6: A frame from an example schematic shown to survey respondents.

The trajectories were rendered as animated gif images, in which the pedestrian was depicted as a small red circle, and vehicles as blue rectangles. A random rotation was added to each image to eliminate visual bias due to the layout of the scene. The spatial limits of the animation were set such that the users had an ample field of view to see oncoming vehicles from both sides (which is important for judging pedestrian behavior). All animations were set to have a similar field of view. For the real-life trajectories, only the single jaywalking pedestrian was annotated. Scenes were chosen such that the pedestrians kept an ample distance from other pedestrians such that they did not visibly interact or influence each other's motion; textiti.e., the vehicle-pedestrian interaction would be the sole interaction evaluated. An example schematic in the style of the animations is shown in Figure 6.

The survey consisted of 3 forced-choice questions featuring a comparison between either a real-life trajectory and a VR-collected trajectory, a real-life trajectory and a synthetic piecewise-linear trajectory, or a VR-collected trajectory and a synthetic piecewise-linear trajectory. Next to each pair, we prompted respondets with the question, *Which one looks more real?* The synthetic piecewise-linear trajectory was included as a control; expectation is that users would more easily perceive that trajectory as "fake." Pairings was selected at random, choosing one animation from 7 real-life trajectories, one from 8 VR trajectories, and one from 2 synthetic trajectories. With 10% probability, the VR trajectory was swapped out for the synthetic trajectory, and with 10% probability, the real-life trajectory was swapped out for the synthetic trajectory; thus, $\sim 20\%$ of pairings featured a synthetic trajectory in the comparison. The other $\sim 80\%$ of pairings featured a comparison between real-life and VR trajectory. Display order was randomized. Finally, there was an optional free-response question at the end: *What characteristics did you use to distinguish between real and fake?* The full set of animations for each of the three categories will be provided in the supplementary material.

The survey was distributed via word-of-mouth as well as by university mailing list. Response time for the survey was around 3 minutes per person. Due to the low time-commitment nature of the survey, no compensation was offered for completion of the survey.

B. Results and Analyses

A total of 302 respondents answered the survey for a total of $302 \times 3 = 906$ forced-choice comparisons made by survey respondents.

Consistent with expectations about the "control" variable, evaluators could easily tell the difference between a simple synthetic policy and a genuine human trajectory. 89.3% of real-life / synthetic comparisons were evaluated in favor of the real-life trajectory looking "more real" than the synthetic. The response was similarly high for the VR / synthetic comparisons, with 88.9% evaluated in favor of the virtual over the synthetic (Table IV).

Also consistent with expectations, the real-life trajectories were the most-frequently chosen as "more real," with approximately 67% of responses choosing it as more "real" in a pair (Table IV). However, VR trajectories do not lag far behind; 59% of responses chose the VR trajectories as "more real" in a pair (Table IV). Of particular note is that 36.5% of real-life / VR comparisons were evaluated in favor of the VR trajectory looking "more real" (Table V); these responses claimed that the VR trajectory looked "more real" than a real trajectory.

Though 36.5% is still a bit away from 50%, which would mean that real-life trajectories are indistinguishable from VR

Truth Value → Guessed value ↓	Real-life	VR	Synthetic
Real	67.0%	58.6%	10.9%
Fake	33.0%	41.4%	89.1%

(a) Percentages

Truth Value → Guessed value ↓	Real-life	VR	Synthetic
Real	547	482	19
Fake	269	340	155
Total	816	822	174

(b) Counts

TABLE IV: Confusion matrix for survey respondent guesses (real / fake) vs. truth category (real-life, VR, synthetic)

Pairing	Real-life / VR	Real-life / Synthetic	VR / Synthetic
% Correct (N)	64.5% (732)	89.3% (84)	88.9% (90)

TABLE V: Percent correct of each pairing. "Correct" is defined as selecting real as more "real" when compared with either VR or synthetic, and selecting VR as more "real" when compared to synthetic.

trajectories, this number still substantiates the claim that VR trajectories can evoke genuine pedestrian responses. First of all, the results confirm that real-life trajectories are less distinguishable from VR trajectories than they are from fully-synthetic trajectories, supporting the claim that the VR system evokes more natural pedestrian responses than simple policies used by synthetic pedestrian simulators [9, 38]. Second, there could be other reasons evaluators are able to tell the difference between real and VR trajectories that have nothing to do with how "real" VR trajectories are. For example, one respondent who evaluated 3 comparisons between real-life and VR trajectories and marked all the real-life ones as more "real" gave the free response answer that "the fake pedestrians seem to cross dangerously close to the vehicles, and sometimes stop in the middle of the road..." Though this respondent labelled all real-life trajectories as "more real" than the corresponding VR ones, their explanation for their rating is not necessarily true. It is generally considered dangerous and not commonly observed that pedestrians "cross dangerously close to the vehicles." As observed in all drone-collected videos, the average jaywalker in real-life is a "safe" jaywalker who only jaywalks when there is sufficient gap in the vehicles. However, sometimes there are pedestrians who do engage in risk-taking behavior. Sometimes there are jaywalkers who jaywalk even on heavily-trafficked roads. These jaywalkers, not able to find a gap between vehicles, must stop in the middle of the road before they can cross completely. Though rare, "dangerous" jaywalkers do exist in real-life, and it is important that safety-critical AV systems have their data.

VII. LIMITATIONS, DISCUSSION, AND FUTURE WORK

A. Interview Responses from Part 1 Not Addressed by VR System

There were some concerns elicited by the participants of the Part 1 structured interview that were not addressed by the evaluated VR pedestrian simulator. Some of these concerns form the foundation for future work:

1) I1 and I2 suggested that the data from vehicle simulators may be less noisy than real-life data, enlarging the sim-to-real gap. In the VR CARLA environment, vehicles follow simple policies like the Intelligent Driver Model, always driving perfectly along predefined paths. This may not realistically reproduce human driving behavior, which is imperfect, and sometimes deviates from the center of the lane. In the future, more complex and noisy driving policies can be implemented into the CARLA environment to create more realistic vehicle behavior. In addition, as described in the concurrent work contributing the VR system [3], a real human driver can be hooked into the simulator as an additional vehicle agent. The real human driver will create vehicle trajectories that pedestrians users may deem more natural.

2) I4 pointed out that certain object categories were under-represented in their company's large internal dataset, thus impacting the ability of the AV system to detect them. For example, small children, specific types of signage such as construction zone signs, skateboarders, trikes, and bicycles in particular were lacking. Future work includes inviting children and elderly to collect data with our system, so we can make use of the low-risk benefits of the VR environment to obtain the lacking but much-needed behavioral data of vulnerable road users.

B. Free Responses from Part 2 Eliciting System Limitations

The qualitative responses summarized in Section V-C2 to the free response question, *What aspects of the system or environment were unrealistic?*, revealed certain limitations of the VR system that can be expanded in future work. While some concerns are difficult to address with the state of current technologies, such as (1), (2), and (4), other concerns can be resolved with additional engineering of the VR system or revised study design. To address (3), giving participants a longer time to familiarize themselves with the VR headset may increase their reassurance that they will not run into obstacles in the real world. To address (5), sound can be easily added to the simulation as it is available in CARLA. However, we chose not to use it in our study, so that participants could hear us when we verbally instructed participants to where they should walk. (6) can be addressed by adding into the scene the relevant environmental elements, all of which are available as assets in Unreal Engine. (7) may be a bit more difficult to address across all environments, but can be addressed for the sidewalk / curb situation by adding physical objects in the real world to match those in the virtual world, and carefully calibrating the virtual-to-real mapping. (8)

can be addressed by integrating into the system the driving simulator extension, which has already been developed. A human driver participant can then use the driving simulator to operate the vehicles in more natural way. Furthermore, the vehicles currently move based on default CARLA policies; better policies can be integrated into CARLA for more realistic vehicle movement. (10) can be addressed by adding in additional pedestrians into the simulation, which the system also supports; each pedestrian would require an additional VR setup, including headset, wireless tracker, and computer. Alternatively, autonomous pedestrians could be added into the simulator. Although these pedestrians would move based on simple policies, these policies are reactive to the user's movements. To address (11), participants can be given tasks to do rather than specific goals to navigate to, such as "Please drop off a letter in the nearest mail box," rather than "Please walk across the road."

With some design modifications and system engineering, nearly all of the concerns raised by the users can be addressed.

C. Part 3 Study Design Choice Limitations

One limitation of the design of our evaluative study is that there may not be enough information from a simple 2D schematic for a human evaluator to determine if a trajectory is "real." As shown by the results, there is enough information present in the 2D trajectory schematics to distinguish between a real-life and a simple synthetic linear policy, as well as between a VR and a simple synthetic linear policy. However, the evaluator has no access to information that could be used to make a better decision, such as body pose. As future work, a body-pose data collection system can be set up in real-life and used to collect body pose data from real-life jaywalkers. The body pose data can be used to create animations for a third-person 3D full-body trajectory comparative evaluation.

Again, it is important to note that just because a VR-collected trajectory is distinguishable from a real trajectory by a human evaluator, that does not necessarily imply that it is unrealistic or unseen in real-life. The VR system specializes in collecting data of uncommon scenarios. This means that even though the VR-collected data may exhibit aggressive behavior, it is behavior that may still be seen in real-life.

VIII. CONCLUSION

Safely and ethically collecting pedestrian trajectory data for training reliable autonomous vehicle models is a complex and nuanced problem; but Virtual Reality data collection offers a solution. A VR data collection system fills the need for diverse and comprehensive data while overcoming the difficulty of obtaining such data in the real world. Through semi-structured interviews, a first-person evaluation, and a third-person survey, we provide substantial evidence that VR-based system can evoke a genuine response from pedestrian users, and thus offers an effective data collection solution.

Despite some limitations in the sense of presence within the virtual environment, the data collected still appears to be reasonably similar to that collected in real-life scenarios.

This presents a compelling case for VR as an effective tool for gathering pedestrian data, particularly in complex, high-risk, or ethically challenging scenarios that are difficult to replicate in the real world. These findings add a new dimension to our understanding of vehicle-pedestrian data collection, emphasizing the potential for VR technology to play a vital role in shaping the future of safe and reliable autonomous transport.

REFERENCES

[1] Mavic mini - DJI. https://www.dji.com/mavic-mini. Accessed: 2023-9-15.

[2] Sander Ackermans, Debargha Dey, Peter Ruijten, Raymond H Cuijpers, and Bastian Pfleging. The effects of explicit intention communication, conspicuous sensors, and pedestrian attitude in interactions with automated vehicles. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA, April 2020. Association for Computing Machinery.

[3] Anonymized. Vehicle-pedestrian interaction data generation for making prediction model robust with vr pedestrian simulator for autonomous driving. *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2023. Under Review.

[4] Silvia Arias, Axel Mossberg, Daniel Nilsson, and Jonathan Wahlqvist. A study on evacuation behavior in physical and virtual reality experiments. *Fire Technol.*, 58(2):817–849, March 2022.

[5] John T Bell and H Scott Fogler. The investigation and application of virtual reality as an educational tool. https://www.cs.uic.edu/~jbell/Professional/Papers/aseepap2.pdf. Accessed: 2023-9-5.

[6] Mehmet Ilker Berkman and Ecehan Akan. Presence and immersion in virtual reality. In Newton Lee, editor, *Encyclopedia of Computer Graphics and Games*, pages 1–10. Springer International Publishing, Cham, 2019.

[7] Florian Berton, Anne-Hélène Olivier, Julien Bruneau, Ludovic Hoyet, and Julien Pettre. Studying gaze behaviour during collision avoidance with a virtual walker: Influence of the virtual reality setup. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 717–725, March 2019.

[8] Rajaram Bhagavathula, Brian Williams, Justin Owens, and Ronald Gibbons. The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.*, 62(1):2056–2060, September 2018.

[9] Abhijat Biswas, Allan Wang, Gustavo Silvera, Aaron Steinfeld, and Henny Admoni. SocNavBench: A grounded simulation testing framework for evaluating social navigation. February 2021.

[10] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qual. Res. Psychol.*, 3(2):77–101, January 2006.

[11] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh

Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. March 2019.

[12] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3D tracking and forecasting with rich maps. November 2019.

[13] Wenxiang Chen, Qianni Jiang, Xiangling Zhuang, and Guojie Ma. Comparison of pedestrians' gap acceptance behavior towards automated and human-driven vehicles. In *Engineering Psychology and Cognitive Ergonomics. Cognition and Design*, Lecture notes in computer science, pages 253–261. Springer International Publishing, Cham, 2020.

[14] Yun Chen, Frieda Rong, Shivam Duggal, Shenlong Wang, Xinchen Yan, Sivabalan Manivasagam, Shangjie Xue, Ersin Yumer, and Raquel Urtasun. GeoSim: Realistic video simulation via Geometry-Aware composition for Self-Driving. January 2021.

[15] Mark Colley and Enrico Rukzio. A design space for external communication of autonomous vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '20, pages 212–222, New York, NY, USA, September 2020. Association for Computing Machinery.

[16] Ana Fiona Dalipi, Dongfang Liu, Xiaolei Guo, Yingjie Chen, and Christos Mousas. VR-PAVIB: The virtual reality Pedestrian-Autonomous vehicle interaction benchmark. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '20, pages 38–41, New York, NY, USA, September 2020. Association for Computing Machinery.

[17] Koen de Clercq, Andre Dietrich, Juan Pablo Núñez Velasco, Joost de Winter, and Riender Happee. External Human-Machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Hum. Factors*, 61 (8):1353–1370, December 2019.

[18] Shuchisnigdha Deb, Daniel W Carruth, Richard Sween, Lesley Strawderman, and Teena M Garrison. Efficacy of virtual reality in pedestrian safety research. *Appl. Ergon.*, 65:449–460, November 2017.

[19] Shuchisnigdha Deb, Daniel W Carruth, and Christopher R Hudson. How communicating features can help pedestrian safety in the presence of self-driving vehicles: Virtual reality experiment. *IEEE Trans. Hum. Mach. Syst.*, 50(2):176–186, April 2020.

[20] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. November 2017.

[21] Tobias Drey, Michael Rietzler, and Enrico Rukzio. Questionnaires and qualitative feedback methods to measure user experience in mixed reality. April 2021.

[22] Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp, Charles Qi, Yin Zhou, Zoey Yang, Aurelien Chouard, Pei Sun, Jiquan Ngiam, Vijay Vasudevan, Alexander McCauley, Jonathon Shlens, and Dragomir Anguelov. Large scale interactive motion forecasting for autonomous driving : The waymo open motion dataset. April 2021.

[23] Ilja T Feldstein and Georg N Dyszak. Road crossing decisions in real and virtual environments: A comparative study on simulator validity. *Accid. Anal. Prev.*, 137: 105356, March 2020.

[24] Yan Feng, Dorine Duives, Winnie Daamen, and Serge Hoogendoorn. Data collection methods for studying pedestrian behaviour: A systematic review. *Build. Environ.*, 187(107329):107329, January 2021.

[25] Tanja Fuest, Anna Sophia Maier, Hanna Bellem, and Klaus Bengler. How should an automated vehicle communicate its intention to a pedestrian? – a virtual reality study. In *Human Systems Engineering and Design II*, pages 195–201. Springer International Publishing, 2020.

[26] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The KITTI dataset. *Int. J. Rob. Res.*, 32(11): 1231–1237, September 2013.

[27] Victoria Gitelman, Sharon Levi, Roby Carmel, Anna Korchatov, and Shalom Hakkert. Exploring patterns of child pedestrian behaviors at urban intersections. *Accid. Anal. Prev.*, 122:36–47, January 2019.

[28] Mar Gonzalez-Franco and Tabitha C Peck. Avatar embodiment. towards a standardized questionnaire. *Front Robot AI*, 5:74, June 2018.

[29] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human mental workload.*, 382: 139–183, 1988.

[30] Tilo Hartmann, Werner Wirth, Holger Schramm, Christoph Klimmt, Peter Vorderer, André Gysbers, Saskia Böcking, Niklas Ravaja, Jari Laarni, Timo Saari, Feliz Gouveia, and Ana Maria Sacau. The spatial presence experience scale (SPES): A short self-report measure for diverse media settings. *Journal of Media Psychology: Theories, Methods, and Applications*, 28(1): 1–15, 2016.

[31] Juan Luis Higuera-Trujillo, Juan López-Tarruella Maldonado, and Carmen Llinares Millán. Psychological and physiological human responses to simulated and real environments: A comparison between photographs, 360° panoramas, and virtual reality. *Appl. Ergon.*, 65:398–409, November 2017.

[32] Shiyu Huang and Deva Ramanan. Expecting the unexpected: Training detectors for unusual pedestrians with adversarial imposters. March 2017.

[33] Shiyu Huang and Deva Ramanan. Expecting the unexpected: Training detectors for unusual pedestrians with adversarial imposters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2243–2252, 2017.

[34] Omar Janeh, Gerd Bruder, Frank Steinicke, Alessandro

Gulberti, and Monika Poetter-Nerger. Analyses of gait parameters of younger and older adults during (Non-)Isometric virtual walking. *IEEE Trans. Vis. Comput. Graph.*, 24(10):2663–2674, October 2018.

[35] Suresh Kumaar Jayaraman, Chandler Creech, Lionel P Robert, Jr, Dawn M Tilbury, X Jessie Yang, Anuj K Pradhan, and Katherine M Tsui. Trust in AV. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, March 2018. ACM.

[36] Matthew Johnson-Roberson, Charles Barto, Rounak Mehta, Sharath Nittur Sridhar, Karl Rosaen, and Ram Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? *arXiv preprint arXiv:1610.01983*, 2016.

[37] Wonhui Kim, Manikandasriram Srinivasan Ramanagopal, Charles Barto, Ming-Yuan Yu, Karl Rosaen, Nick Goumas, Ram Vasudevan, and Matthew Johnson-Roberson. PedX: Benchmark dataset for metric 3D pose estimation of pedestrians in complex urban intersections. September 2018.

[38] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Trans. Intell. Transp. Syst.*, 23(7):7386–7400, July 2022.

[39] Yee Mun Lee, Ruth Madigan, Jorge Garcia, Andrew Tomlinson, Albert Solernou, Richard Romano, Gustav Markkula, Natasha Merat, and Jim Uttley. Understanding the messages conveyed by automated vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '19, pages 134–143, New York, NY, USA, September 2019. Association for Computing Machinery.

[40] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. *Comput. Graph. Forum*, 26(3):655–664, September 2007.

[41] Lan Li, Fei Yu, Dongquan Shi, Jianping Shi, Zongjun Tian, Jiquan Yang, Xingsong Wang, and Qing Jiang. Application of virtual reality technology in clinical medicine. *Am. J. Transl. Res.*, 9(9):3867–3880, September 2017.

[42] Junwei Liang, Lu Jiang, Kevin Murphy, Ting Yu, and Alexander Hauptmann. The garden of forking paths: Towards Multi-Future trajectory prediction. December 2019.

[43] Karthik Mahadevan, Elaheh Sanoubari, Sowmya Somanath, James E Young, and Ehud Sharlin. AV-Pedestrian interaction design using a pedestrian mixed traffic simulator. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, pages 475–486, New York, NY, USA, June 2019. Association for Computing Machinery.

[44] R McCloy and R Stone. Science, medicine, and the future. virtual reality in surgery. *BMJ*, 323(7318):912–915, October 2001.

[45] Leonel Merino, Magdalena Schwarzl, Matthias Kraus, Michael Sedlmair, Dieter Schmalstieg, and Daniel Weiskopf. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). October 2020.

[46] David Michalík, Miroslav Jirgl, Jakub Arm, and Petr Fiedler. Developing an unreal engine 4-based vehicle driving simulator applicable in driver behavior Analysis—A technical perspective. *Safety*, 7(2):25, April 2021.

[47] Alexandre M Nascimento, Anna Carolina M Queiroz, Lucio F Vismari, Jeremy N Bailenson, Paulo S Cugnasca, João B Camargo Junior, and Jorge R de Almeida. The role of virtual reality in autonomous vehicles' safety. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 50–507, December 2019.

[48] Niels Christian Nilsson, Stefania Serafin, and Rolf Nordahl. The effect of head mounted display weight and locomotion method on the perceived naturalness of virtual walking speeds. In *2015 IEEE Virtual Reality (VR)*, pages 249–250, March 2015.

[49] Niels Christian Nilsson, Tabitha Peck, Gerd Bruder, Eri Hodgson, Stefania Serafin, Mary Whitton, Frank Steinicke, and Evan Suma Rosenberg. 15 years of research on redirected walking in immersive virtual environments. *IEEE Comput. Graph. Appl.*, 38(2):44–56, March 2018.

[50] Nahal Norouzi, Kangsoo Kim, Myungho Lee, Ryan Schubert, Austin Erickson, Jeremy Bailenson, Gerd Bruder, and Greg Welch. Walking your virtual dog: Analysis of awareness and proxemics with simulated support animals in augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 157–168, October 2019.

[51] J Pablo Nuñez Velasco, Haneen Farah, Bart van Arem, and Marjan P Hagenzieker. Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality. *Transp. Res. Part F Traffic Psychol. Behav.*, 66:1–14, October 2019.

[52] Office for Human Research Protections (OHRP). Special protections for children as research subjects. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/special-protections-for-children/index.html, November 2010. Accessed: 2023-9-10.

[53] Nami Ogawa, Takuji Narumi, Hideaki Kuzuoka, and Michitaka Hirose. Do you feel like passing through walls?: Effect of Self-Avatar appearance on facilitating realistic behavior in virtual environments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, New York, NY, USA, April 2020. Association for Computing Machinery.

[54] Anne-Helene Olivier, Julien Bruneau, Richard Kulpa, and Julien Pettre. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments. *IEEE Trans. Vis. Comput. Graph.*, 24(7):2251–2263, July 2018.

[55] S Pellegrini, A Ess, K Schindler, and L van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE, September 2009.

[56] Iana Podkosova and Hannes Kaufmann. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, number Article 21 in VRST '18, pages 1–11, New York, NY, USA, November 2018. Association for Computing Machinery.

[57] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John Tsotsos. PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, October 2019.

[58] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Computer Vision – ECCV 2016*, pages 549–565. Springer International Publishing, 2016.

[59] Gayani Samaraweera, Rongkai Guo, and John Quarles. Latency and avatars in virtual environments and the effects on gait for persons with mobility impairments. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 23–30, March 2013.

[60] Sonja Schneider and Klaus Bengler. Virtually the same? analysing pedestrian behaviour by means of virtual reality. *Transp. Res. Part F Traffic Psychol. Behav.*, 68: 231–256, January 2020.

[61] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. Stanford sleepiness scale (SSS). In Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro, editors, *STOP, THAT and One Hundred Other Sleep Scales*, pages 369–370. Springer New York, New York, NY, 2012.

[62] Gustavo Silvera, Abhijat Biswas, and Henny Admoni. DReyeVR: Democratizing virtual reality driving simulation for behavioural & interaction research. January 2022.

[63] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments. *Presence: Teleoper. Virtual Environ.*, 6(6):603–616, December 1997.

[64] Kay M Stanney, Mansooreh Mollaghasemi, Leah Reeves, Robert Breaux, and David A Graeber. Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. *Int. J. Hum.-Comput. Stud.*, 58(4):447–481, April 2003.

[65] Chek Tien Tan, Leon Cewei Foo, Adriel Yeo, Jeannie Su Ann Lee, Edmund Wan, Xiao-Feng Kenan Kok, and Megani Rajendran. Understanding user experiences across VR walking-in-place locomotion methods. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, number Article 517 in CHI '22, pages 1–13, New York, NY, USA, April 2022.

Association for Computing Machinery.

[66] C X O Tina Øvad. [no title]. https://preely.com/likert-scale-semantic-differential-scale/. Accessed: 2023-8-29.

[67] Tram Thi Minh Tran, Callum Parker, and Martin Tomitsch. A review of virtual reality studies on autonomous Vehicle–Pedestrian interaction. *IEEE Transactions on Human-Machine Systems*, 51(6):641–652, December 2021.

[68] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence*, 9(5):497–503, October 2000.

[69] Florian Weber, Ronee Chadowitz, Kathrin Schmidt, Julia Messerschmidt, and Tanja Fuest. Crossing the street across the globe: A study on the effects of eHMI on pedestrians in the US, germany and china. In *HCI in Mobility, Transport, and Automotive Systems*, pages 515–530. Springer International Publishing, 2019.

[70] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next generation datasets for Self-Driving perception and forecasting. January 2023.

[71] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, June 1998.

[72] Ze Yang, Yun Chen, Jingkang Wang, Sivabalan Manivasagam, Wei-Chiu Ma, Anqi Joyce Yang, and Raquel Urtasun. UniSim: A neural closed-loop sensor simulator. August 2023.

[73] Katja Zibrek and Rachel McDonnell. Social presence and place illusion are affected by photorealism in embodied VR. In *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*, number Article 13 in MIG '19, pages 1–7, New York, NY, USA, October 2019. Association for Computing Machinery.