Developing a VR-Integrated Framework to Enhance Agent-Based Model in Shared Space through Behavioral Data Collection and Analysis

Master Thesis Geo-Information Science

Developing a VR-integrated Framework to Enhance Agent-based Model in Shared Space through Behavioral Data Collection and Analysis

Shuying Guo

1051297

Course number: GRS80436

Supervisor: Jiayan Zhao PhD

Co-supervisor: dr.ir. Arend Ligtenberg

Date of delivery: 13/08/2025

Thesis submitted in partial fulfilment of the requirements for the Master of Science degree in Geo-Information Science - Wageningen University & Research (WUR) with support from WANDER Lab.

Acknowledgement

This thesis would not have been possible without the support of many wonderful people.

I am deeply grateful to my supervisors, Jiayan and Arend, for their guidance, patience, and continuous support.

Mingyang encouraged me through difficult moments and helped me believe in myself.

Gaku, even from far away, offered steady support that helped me get through lame days.

Dongbo gave me invaluable help and taught me how to manage my time more wisely.

And all the people I met at WUR, thank you for accompanying me with the long, challenging, meaningful, and enjoyable journey.

Life is good!

Abstract

Urbanization has intensified traffic in constrained urban spaces, leading to the adoption of shared spaces, which pose safety challenges due to complex multi-user interactions. Agent-based models (ABMs) can simulate these interactions but often lack empirical data for accurate calibration. Virtual Reality (VR) offers a powerful platform for collecting behavioral data to refine ABMs, yet most VR-based experiments focus on a single road-user type in simplified environments, limiting their applicability to complex shared spaces. Furthermore, experimental setups are often designed only for data collection, overlooking how the design itself may influence results and requiring additional effort for analysis and redesign. This study addresses these gaps by developing a VR-integrated ABM framework for collecting behavioral data under configurable environmental conditions, with results visualized and analyzed via a Streamlit dashboard. The framework was adapted from Linnekamp's ABM, which simulates traffic conflicts in a shared space near Amsterdam Central Station by calculating conflict probabilities based on agent interactions. This study extended this ABM to incorporate two conflict probability types - physical (based on the 3D model boundary) and visual (based on human perceptual boundaries) - and developed a user interface to collect user-perceived conflicts for comparison with system-calculated values. All behavioral and system performance data, including frame rate (FPS) and latency, were stored in structured .csv files and visualized through the dashboard. A small-scale experiment with ten participants evaluated the framework's Quality of Experience (QoE) and analyzed behavioral patterns to identify opportunities for refinement. Participants reported high presence and usability, with moderate cybersickness; user-suggested improvements included multisensory feedback and richer environmental variation. Low-density scenarios achieved higher median FPS than high-density ones, with similar latency. Behavioral analysis revealed varied avoidance strategies: high density led to more frequent pauses and lower speeds, user-marked conflicts often diverged from system-calculated probabilities, and cyclists were more often perceived as conflict sources in low density, pedestrians in high density. Limitations included: (1) implementation issues such as logging misalignment, delayed UI feedback, and insufficient visual distinction between agent types; (2) QoE constraints including unrealistic agent motion, limited sensory cues, and cybersickness linked to low FPS; (3) small, imbalanced samples favoring pedestrians and high-density scenarios. Future work should optimize functionality, more diversify scenarios and participant profiles, and expand applications for ABM calibration across broader mobility contexts. While the framework is still preliminary and limited in scope, it offers a practical starting point for developing more efficient and empirically grounded ABMs. The proposed approach and findings lay a foundation for advancing traffic modeling, simulation-based evaluation, and human-centered transport system design.

Key Words: Agent-Based Modeling (ABM); Shared Space; Traffic Simulation; Human Behavior

Contents

1. Introduction

Urbanization has led to an increase in road users, while urban road space remains limited. As a result, pedestrians, cyclists, and motorized vehicles are often required to share the same constrained spaces, leading to more frequent interactions and a higher risk of conflicts. Such mixed-traffic settings without clear separation or priority systems, have become a common feature of urban traffic environment (Allocated & Core, 2013; Drut, 2018).

In response, the Shared Space concept emerged, aiming to replace formal controls with informal negotiation, encouraging mutual awareness and self-organizing interactions (Hamilton-Baillie, 2008). Over the past decades, Shared Space has been increasingly implemented, particularly in Northern Europe. While this design approach removes strict lane demarcations and traffic controls to promote mutual awareness and more cautious behavior, it also introduces unpredictability, raising concerns about safety and user comfort (Kenworthy & Newman, 2015). Empirical findings on its effectiveness remain mixed. Some cases report reduced accidents (e.g., Netherlands, Denmark, UK, Germany) (Edquist & Corben, 2012; Methorst et al., 2007), while others, like Halifax (UK) and Oosterwolde (NL), show increased crash rates or reduced perceived safety (Edquist & Corben, 2012; Moody & Melia, 2014a). These mixed findings suggest that the success of shared spaces is highly context-dependent (Quimby & Castle, 2006), influenced by factors such as traffic volume (Moody & Melia, 2014b) and spatial configuration (Batista & Friedrich, 2022). Among these, the way road users interact and adapt to each other plays a particularly critical role, as shared spaces rely less on formal controls and more on informal social negotiation (Kaparias et al., 2015).

Therefore, evaluating the performance of shared spaces is essential to understand how different users interact with others, and to identify potential safety risks, behavioral inefficiencies, and unintended consequences for planning safe traffic environment. To do so, micro traffic simulation, also known as microscopic traffic modeling, has been an efficient approach. This modelling does simulations based on behavior and interactions of road users at a high spatial and temporal resolution (Anvari et al., 2015; Barceló, 2010; Burghout et al., 2005). In the context of shared spaces, where the safety result is mainly dominated by road users' interactions, micro traffic simulation is particularly useful, supporting researchers to analyze how road users adapt their behavior in real-time interactions, and to assess how these individual responses influence safety outcomes and overall performance.

1.1 Agent-Based Model (ABM)

Agent-Based Model (ABM) has emerged as a powerful method for simulating road user behaviors in the field of micro traffic simulation (Ljubović, 2009). In ABM, each road user is represented as an autonomous agent capable of perceiving the environment and making decisions based on a series components, including agent attributes, behavioral rules, interaction modelling mechanisms, and environmental cues (Bonabeau, 2002; Macal & North, 2005a, 2005b). Agent attributes define the internal states or fixed characteristics of each agent. These can include physical properties (e.g. walking speed, radius, or field of view, etc.), cognitive traits (e.g. reaction time, comfort distance, etc.), as well as social-demographic characteristics (e.g. gender, age, education level, etc.) (Basak et al., 2013; Macal & North, 2005b; Shaaban & Abdelwarith, 2020). Behavioral rules define when an agent should react to perceived stimuli, such as decelerating upon detecting a nearby obstacle. Environmental cues provide the spatial and situational context in which agents operate, including road networks, obstacle maps, traffic signals, and dynamic agent appearances (Bazghandi, 2012).

When building up an ABM, one of the most challenging tasks is designing appropriate behavioral rules and interaction mechanisms. Current studies usually resort to existing behavioral theories, such as social force model (Helbing & Molnar, 1995) and heuristic algorithm (Moussaïd et al., 2011). In recent years, some researchers have begun exploring data-driven approaches, including deep learning techniques (Johora et al., 2020), but these methods often fail to accurately estimate agent-specific micro-level variables and heavily depend on training data (Monti et al., 2023). Among these approaches, the Social Force Model is generally regarded as one of the most effective frameworks for behavior analysis, due to its high interpretability and broad applicability across different traffic contexts (Kouskoulis et al., 2018). To better illustrate the components of ABM and how different modeling approaches align with them, an overview is shown in Fig 1.

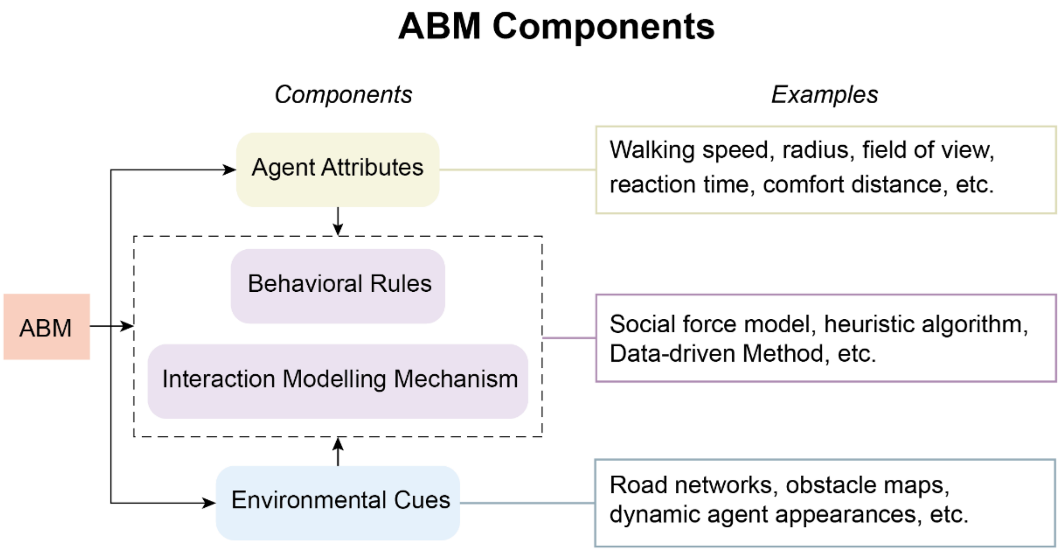

Fig 1. Agent-based Model (ABM) Components

The social force model defines behavioral rules and interaction mechanisms based on a series of agent attributes related to speed, such as desired speed, maximum speed, and field of view, which influence individual mobility. The behavioral rules are in the form of attractive and repulsive effects: agents are drawn toward goals while avoiding others or obstacles. These behaviors are implemented through a continuous interaction mechanism based on force dynamics, where the acceleration or deceleration of each agent depends on the presence and movement of nearby agents within their field of view. In practice, recent studies have modified this theory by adjusting parameters or incorporating new agent types to reflect diverse traffic contexts. For instance, Huang, et al. (2023) and Dias, et al. (2018) applied the social force model as the behavioral foundation of their ABMs to simulate interactions between pedestrians and cyclists or Segway riders in China and Japan, respectively.

To build ABMs with high efficiency and realism, proper calibration and validation are essential for fine-tuning. Calibration helps align model outcomes with real-world behaviors by adjusting agent attributes and decision logic until simulated results match empirical observations (Bianchi et al., 2007). However, this session still faces limitations. The major challenge lies in the lack of high-quality behavioral data to calibrate and validate agent rules, especially in context-sensitive environments like shared spaces. As Bonabeau (2002) and Macal & North (2005a) point out, agent behavior in many ABMs are based on simplified assumptions or theoretical constructs rather than empirically observed decision-making patterns. But human behavior is deeply influenced by contextual and sensory cues, including spatial layout, perceived risk, and environmental familiarity (Angioi & Bassani, 2022; D'Haese et al., 2014; Krumm, 2010; Zhu et al., 2022). This mismatch can lead to biased or unrealistic simulations. To deal with this drawback, one potential improvement to the agent-based model approach is to incorporate real behavioral data, using empirically observed patterns to inform and refine ABMs (J. Huang et al., 2022).

1.2 Virtual Reality for ABM

Virtual Reality (VR) has become a promising tool for collecting behavioral data in traffic simulation. Compared to traditional methods like questionnaires and field studies, VR has several advantages: it can capture real behavior, control experimental conditions, and repeat tests easily (Pan & Hamilton, 2018). While desktop-based applications can also simulate virtual environments under controlled settings, VR provides higher immersion through a broader field of view and natural, body-based interaction enabled by 3D-in-3D displays (e.g., head tracking, stereoscopy), leading to a stronger sense of presence, which makes people behave more naturally behavior and allows more convincing insights for behavioral studies in traffic simulation domain (Simpson, 2020; Zhao et al., 2020). Multiple studies have demonstrated that high-immersion VR can elicit realistic user behavior, showing strong alignment with real-world actions across various traffic scenarios, including pedestrian movement (Bhagavathula et al., 2018), crossing decisions (Angulo et al., 2024), and driving behavior such as speed control and steering (Blissing et al., 2019; Piaseczna et al., 2024).

When developing a VR application for micro traffic simulation, to consistently achieve such realism, Quality of Experience (QoE) offers a comprehensive framework to evaluate a VR system's usability and immersion level. It includes subjective factors such as presence (e.g. presence/immersion, usability, cybersickness, emotion) (Slater, 2009) and objective factors related with system performance (e.g. frame rate, latency, resolution) (Gaggioli et al., 2003). High QoE generally encourages more natural behavior, providing valuable insights for transport simulation (Chamilothori et al., 2019; Slater, 2009). However, if users perceive agent movements as unrealistic (e.g. low presence/immersion) or the interface as laggy (e.g., low FPS or high latency), they may not respond naturally, leading to distorted behavioral data. Therefore, assessing QoE has become essential in VR-based traffic research, as it helps determine whether the application can reliably elicit realistic behaviors in virtual environments (Neo et al., 2021).

To implement such immersive, behaviorally grounded simulations, Unity has emerged as a powerful development platform. As a game engine, Unity supports real-time interaction, rich 3D visualization, and flexible user interface design, making it ideal for VR-integrated agent-based modeling (ABM). For instance, Cheliotis (2022) modeled the common indoor crowd behavior by a Unity framework of ABM (ABMU), and Huang, et al. (2023) developed a Unity-BIM application to implement evacuation ABM. Unlike traditional platforms for building ABM-based traffic simulations, such as MATSim (Balmer et al., 2008), GAMA (Taillandier et al., 2019), and NetLogo (Wilensky & Rand, 2015), which primarily focus on abstract 2D simulations and rarely incorporate human perception, Unity allows for rich 3D visualization and immersive interaction, making it especially suitable for stimulating and collecting human behavior in complex environments (Brookes et al., 2020). With C# as its primary scripting language, Unity supports high customizability in defining agent logic, environmental responses, and user input. In addition, Unity supports the creation of highly realistic and visually rich environments, which further enhance the QoE level. More importantly, it enables real-time collection of behavioral data during simulation, which is essential for ABM refining and studying decision-making processes.

1.3 Behavioral Features of Traffic Agents

In micro-traffic simulations, human behavioral features, which refer to the physical behavior responding to the environment (e.g. speed, direction, etc.), play a central role in modeling and predicting individual decisions and movement dynamics. However, as highlighted in the previous section, many agent-based models remain overly simplified and often lack the incorporation of human behavior. Therefore, it is essential to define which behavioral features are most relevant to micro-traffic contexts and can be feasibly captured within a VR environment.

There is multiple road users involved in micro-traffic simulation, such as pedestrians, drivers, cyclists and mopeds, but studies with a systematic structure on behavior and task modeling mainly revolve around pedestrians and drivers. First pointed out by Michon (1985), drivers' road tasks can be divided into three layers: strategic level, maneuvering level, and control level - Strategic level defines the general planning stage of a trip (e.g. determination of trip, traffic motive); Maneuvering level includes the general behavior towards the destination (e.g. avoidance method, gap acceptance, and overtaking); Control level refers to more specific interactions, like local interaction with spaces and pedestrians (e.g. brake, accelerate and decelerate). Converted from this framework, Hoogendoorn & Bovy (2004) raised a framework categorizing pedestrian behaviors as strategic level (e.g. determination of trip, overall route, travel time), tactical level (e.g. reselect the route, avoidance method, judge crossing time), and operational level (e.g. real-time avoidance, acceleration and deceleration, fine-tuning of walking direction), respectively corresponding to Michon's three layers. For cyclists and moped users, systematic research on behavior is limited, and most existing studies come from practical or applied projects.

In recent practices collecting behavioral data for micro traffic simulation, the number of research about pedestrians and drivers highly exceeds other road user types. For pedestrians, commonly tracked features include movement speed, path or trajectory, acceleration, and proximity to obstacles (Angulo et al., 2024; Guo et al., 2022; Mukoya et al., 2024). Some studies also include hesitation-related indicators such as low-speed pauses ("gap stop"), which may reflect momentary uncertainty or decision-making processes (Jay et al., 2020). These metrics are usually recorded through VR joystick control or desktop input. In studies with eye-tracking equipment, physical responses such as gaze direction and fixation are also captured (Sulle et al., 2025). For drivers, simulation tools often log speed, steering angle, and route choice automatically, some studies also combine VR with physiological sensors to collect data like heart rate and gaze (Michalík et al., 2021; Silvera et al., 2022). Apart from speed, acceleration, path or trajectory, studies about cyclists' behavior also care about pedal movement, braking, and steering angle (Alexander, 2015; Nazemi et al., 2018; Yuan et al., 2018). As to mopeds, there is little study about collecting their behaviors through experiments, current studies still prefer using questionnaires to get psychological insights of this type of road user (Aguilera-García et al., 2021). In general, while a variety of behavioral features have been captured across user types through different tools, VR has increasingly become the preferred platform due to its ability to combine high immersion, flexible interaction, and real-time data logging.

1.4 Environmental Factors in Micro-traffic Simulation

It has long been recognized that environmental conditions significantly influence road user behaviors. These environmental factors can be broadly categorized into several types: (1) Transport conditions (e.g., traffic density, vehicle flow rate, and presence of mixed traffic) (Ayres & Mehmood, 2009; Bella, 2011; Kamal & Farooq, 2023); (2) Spatial layout (e.g., road or lane width, presence of obstacles, and intersection type) (Garcia et al., 2015; J. Huang et al., 2022; Khademi et al., 2024; Soares et al., 2021) (3) Visual elements (e.g., lighting, obstruction of view, signage visibility) (Garcia et al., 2015; Motamedi et al., 2017); (4) Road conditions (e.g., pavement quality, potholes, wet or icy roads) (Moreno et al., 2023); (5) Weather conditions (e.g., rain, snow, fog, temperature, wind) (Chen et al., 2019; Liang et al., 2020); (6) Temporal factors (e.g., peak vs. off-peak hours, day/night) (Chen et al., 2019); and (7) Other contextual elements (e.g., urban/rural setting, perceived safety) (Chen et al., 2019; Khademi et al., 2024). Therefore, it is necessary to incorporate environmental changes when collecting road users' behavioral data.

To collect data under different environmental settings, various platforms and tools have been adapted, which can be simplified into three types: transport simulator, VR, and in-field observation methods. Simulator-based experiments offer highly controllable experiment settings, allowing for specific manipulation of transport features. For instance, Bella and Silvestri (2015) examined driver deceleration behavior in response to obstacles, while

Methorst et al. (2007) used a cycling simulator to assess how different road widths influenced cyclists' maneuvering. In contrast, in-field studies can capture behavior in real-world contexts. Garcia et al. (2015) analyzed cyclist behavior under different lane widths through direct observation, Silvera et al. (2022) used sensors to relate spatial layouts and traffic density to driver behavior, and Yuan et al. (2018) examined cyclist trajectory and steering across varying spatial layouts. In recent years, VR-based methods have gained attention for their ability to simulate complex environments while maintaining high experimental control and low overall experiment cost (Pan & Hamilton, 2018). Compared with traditional transport simulators, VR has advantages in both immersion and adaptability. Studies such as Guo et al. (2022) and Nazemi et al. (2018) used VR to study how changes in traffic density or cycling infrastructure affect the behavior of cyclists and pedestrians, while Khademi et al. (2024) explored how intersection design and visibility conditions influence pedestrian crossing decisions and perceived safety. To summarize, using VR can be a promising method to integrate different environmental factors for road users' behavior collection.

1.5 Research Gaps, Objectives, and Questions

Shared space, as an emerging planning strategy in urban mobility, has been increasingly applied in cities. However, despite its growing real-world practice, shared space remains under-researched, especially in behavioral modeling domains. Current studies have primarily focused on in-field observations, which mainly capture binary interactions such as those between pedestrians and motor vehicles (e.g., yielding, stopping) (Kaparias et al., 2015; Tzouras et al., 2023). But these studies fail in assessing individual-level behaviors such as decision-making timing, conflict perception, and responses to multiple road users. As a result, the complex and dynamic interaction mechanisms in shared space are often oversimplified or overlooked.

On the modelling side, although ABMs have shown useful in modeling micro-traffic systems, most existing frameworks remain theory-driven, which agent decision rules are often based on simplified heuristics or pre-defined parameters, limiting their realism, especially in complex and context-sensitive environments like shared spaces (Bonabeau, 2002; Moussaïd et al., 2011). With VR's strength in allowing people to behave naturally and in a controlled manner, using this technology to collect human behavioral data could be a promising way for ABM calibration. While many VR studies focus on behavioral data collection for improved micro-traffic simulation, they often focus on one specific road user or one kind of interaction (e.g. interaction between pedestrian and driver), which are not useful for the case of shared space where interactions happen among various road users. This methodological separation restricts the ability to understand how road users truly adapt under dynamic conditions.

Moreover, in the case of behavioral data collection, especially in VR-based human experiments, experimental setup plays a critical role in influencing outcomes. Prior research has shown that variations in task design, environment layout, and participant instructions can lead to significantly different behavioral responses, indicating the importance of pre-experiment tools to verify the feasibility and sensitivity of the setup (Aguilar et al., 2024). However, most existing VR-based applications lack real-time feedback mechanisms that would support this kind of iterative design. Instead, they primarily function as data collection tools, requiring researchers to export and analyze raw data externally, often after the experiment has concluded. This post-hoc workflow makes it difficult to identify whether a simulation is producing valid or realistic behavior during runtime, thus breaking the feedback loop needed for scenario adjustment and experimental refinement. To deal with this problem, research in biological agent-based modeling has already proven immediate visualizations (e.g. agent distribution, statistical plots) are an effective solution by allowing researchers to identify preliminary model and experiment setting problems, contributing to identify model and experiment problems at early stage (Seekhao et al., 2016). While in the traffic behavior domain, few systems have been designed to offer researcher-facing feedback dashboards that visualize both behavioral outcomes and ABM system performance metrics. As a result, researchers lack the tools to evaluate how well a specific experimental run reflects realistic interaction patterns or computational feasibility under varying conditions.

Therefore, to address these gaps, this study aims to develop a researcher-facing, VR-integrated ABM framework to support road users' behavioral data collection with a visualization dashboard function where the collected data is analyzed for ABM refining. To operationalize this framework, a case study will be carried out using Linnekamp's ABM (Linnekamp, 2020a), an existing ABM based on social force model which focused on the interactions between multiple road users (pedestrians, cyclists, and mopeds) in a typical shared space in Amsterdam.

Specifically, there are three research objectives:
(1) Design and implement a VR-integrated ABM framework comprising three components: capturing behavioral data in a shared space setting, incorporating variations in environmental factors, and supporting a researcher-facing dashboard that visualizes both behavioral outputs and ABM system performance metrics;
(2) Evaluate the QoE of this framework using standard immersion indicators;
(3) Analyze behavioral insights derived from an experiment with this framework, and compare them with Linnekamp's original ABM to identify opportunities for refinement.
To realize these objectives, the main research question was:

How can a VR-integrated ABM framework be developed, focusing on human behavioral data collection and analysis?

Which can be divided by the following sub-research questions (SRQ):

1. How can behavioral data be captured and linked to the VR-integrated framework?
2. How can environmental factors be incorporated into the VR-integrated framework?
3. How can visualized feedback be added to the VR-integrated framework?
4. How is the QoE of the VR-based ABM framework?
5. What behavioral insights can be derived from the VR-integrated framework across different environmental conditions and road user types, and how to refine Linnekamps' ABM?

The first three sub-research questions focus on the development of a VR-integrated ABM framework: SRQ1 explores how behavioral data can be captured and structured in real time; SRQ2 investigates how environmental factors like traffic density can be operationalized within the simulation; and SRQ3 examines how visualized feedback can support post-experiment analysis. SRQ4 evaluates the quality of user experience (QoE) by combining system performance data and participant feedback. SRQ5 uses the collected behavioral data to reflect on existing ABM rules and propose refinements based on perception-driven and context-aware strategies. All together, these research questions can help address the knowledge gaps and contribute a more adaptable VR-integrated framework in micro mixed-traffic simulations for ABM refinement and future mobility research.

2. Methodology

2.1 Study Area

This study chose a shared space in Amsterdam as the case study site. This site is near the ferry dock at Amsterdam Central Station, a high-traffic area characterized by a mix of pedestrians, cyclists, and mopeds (see Fig 2). This location serves as a vital transportation hub where people arrive by train and transfer to ferries across the IJ River. Pedestrians typically flow between the ferry docks and station entrances, while cyclists and mopeds move both parallel and perpendicular to these pedestrian flows, creating a complex mix of road users with varying speeds and movement patterns. The lack of traffic signals or dedicated lanes encourages a fluid but unpredictable interaction between these groups, making it an ideal site to study the dynamics and safety challenges of shared space.

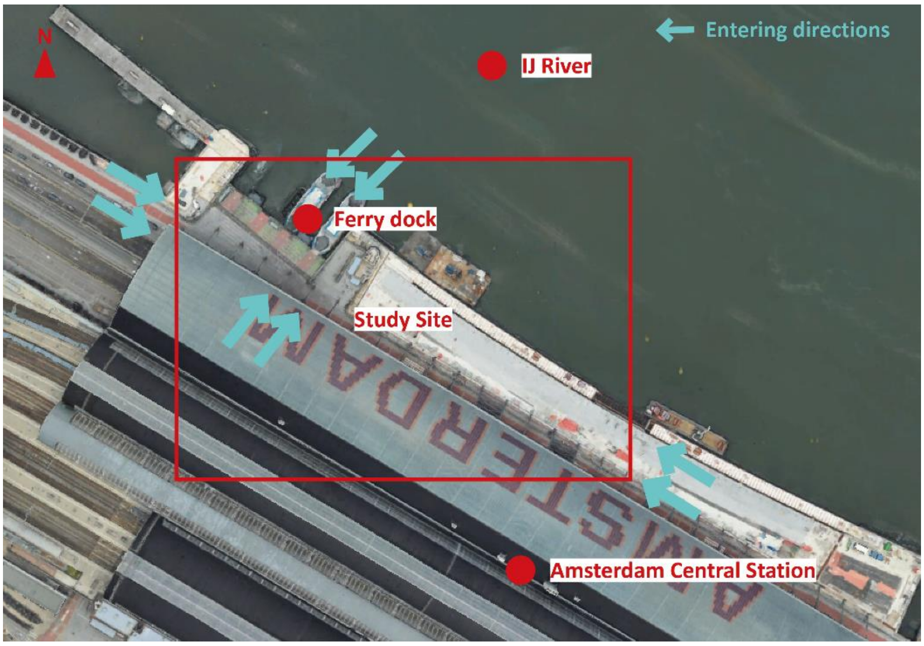

Fig 2. Site Location: Amsterdam Central Station, Amsterdam, the Netherlands (Basemap source: https://www.google.com/maps)

To embed ABM, this study built upon an existing ABM developed by Linnekamp (2020b), which simulated road users' conflicts in a shared space of Amsterdam, the Netherlands. This ABM defined agent attributes including physical size, desired velocity, relaxation time, and reaction time, based on three agent types: pedestrians, cyclists, and mopeds. Agents were generated using pre-defined spawn rates derived from in-field observations and transport reports. As to behavioral rules, it adapted principles from social force model, where agents adjusted their movement in response to surrounding stimuli, such as nearby agents within their field of view. The interaction mechanism involved calculating a conflict probability based on variables such as relative speed, heading direction, edge distance, and local agent density. This conflict probability served as a core output to investigate potential safety risks in the shared space.

However, the conflict probability in Linnekamp's ABM is entirely derived from theory-driven behavioral rules, which may not fully reflect how individuals actually perceive and respond to interactions in real-world shared spaces. In practice, users may adopt different avoidance strategies or interpret situations as conflicts even when no physical collision happens. Such perceived conflicts may still influence user behavior, prompting hesitation, yielding, or path adjustment. Therefore, collecting human behavioral data, including subjective perceptions of conflict, offers valuable input for refining the ABM's behavioral rules and interaction mechanisms, making this ABM an ideal testbed for this study.

2.2 Development of VR-integrated Framework

To conduct this study, there were three main stages: (1) Replicating Linnekamp's ABM in Unity, including agent attributes, behavioral rules, and interaction mechanism; (2) Developing core modules into ABM, including behavioral data collection, environmental condition, and a dashboard for visualization and analysis; (3) Conducting experiments to get user experiences and preliminary insights to improve Linnekamp's ABM. In short, the first two stages were about framework development, and the last was to get user feedback. The architecture of the VR-integrated framework is shown in Fig 3, and the complete implementation code is available in the corresponding GitHub repository as in Appendix 1. Detailed methodology is explained in the following sections.

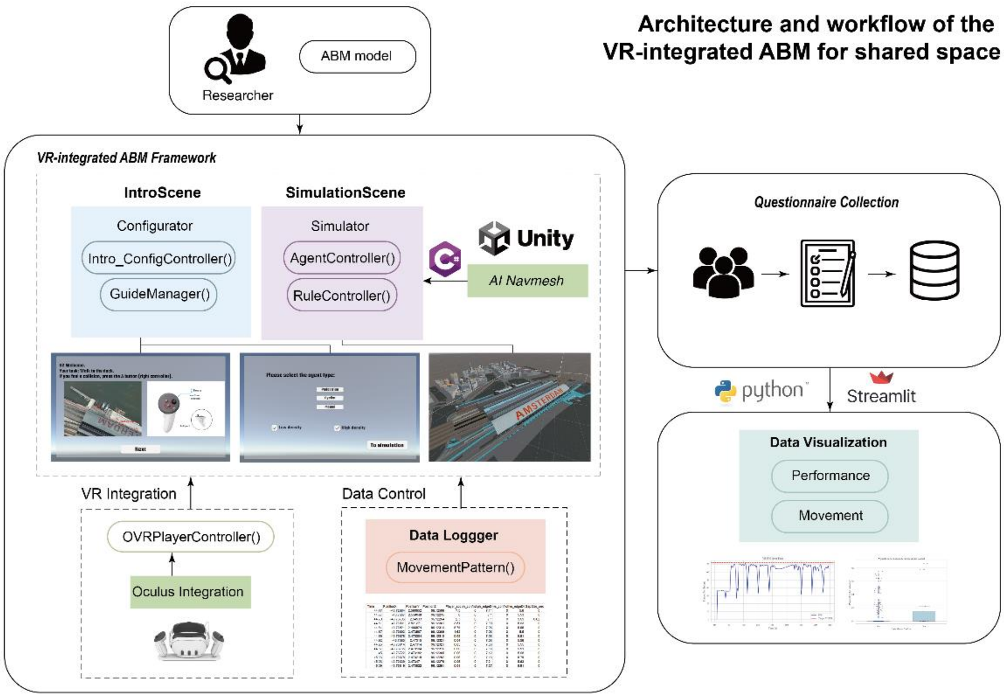

Fig 3. Architecture and workflow of the VR-integrated ABM for shared space

2.2.1 VR Set up and Scenario Configuration

To develop this VR-integrated behavioral simulation framework, the study adopted Unity version 2020.3.40f1 (LTS) for its long-term stability and compatibility with two key packages: AI Navigation and Oculus Integration. The user interface was implemented using Unity's UI Toolkit. A 3D model of Amsterdam Central Station and its surrounding environment was acquired from CGTrader (https://www.cgtrader.com/) to ensure high spatial realism. Agent and environment assets were imported from the "Urban Traffic System Full Pack" of Unity Asset Store (https://assetstore.unity.com), with multiple models used for each agent type to enhance visual diversity (see Fig 4). The simulation was conducted from a first-person perspective using the Oculus XR Plugin 4.2.0, which includes pre-configured tools and scripts to streamline VR development. Although Oculus XR Integration has been recently deprecated, it remains widely used in practice. Its stability, ready-to-use components, and compatibility with Unity's legacy XR framework, makes it easier to build and test interactions during early prototyping, which was important given the limited development timeframe.

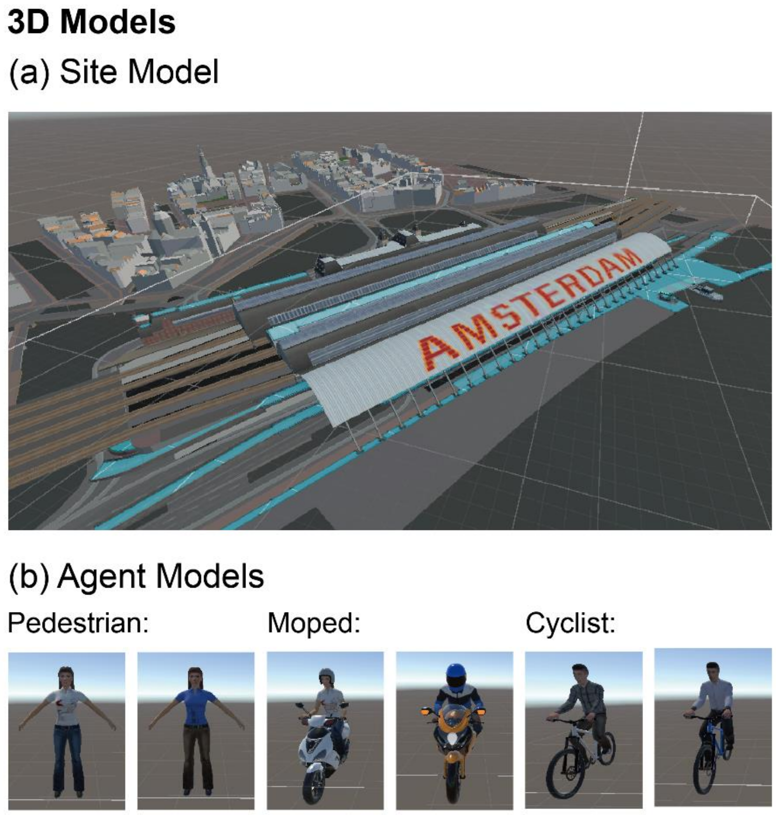

Fig 4. 3D models used in this study. (a) Site model; (b) Agent models.

The simulation was divided into two phases: Intro Scene and Simulation Scene. Before going to the Simulation Scene, participants entered an Intro Scene, where they were introduced to the main task - moving from a designated spawn point to the ferry dock - and familiarized with basic VR operations (see Fig 5). Once the tutorial was complete, the participant was required to select an agent type (pedestrian, cyclist, or moped) and choose between two levels of traffic density (high or low).

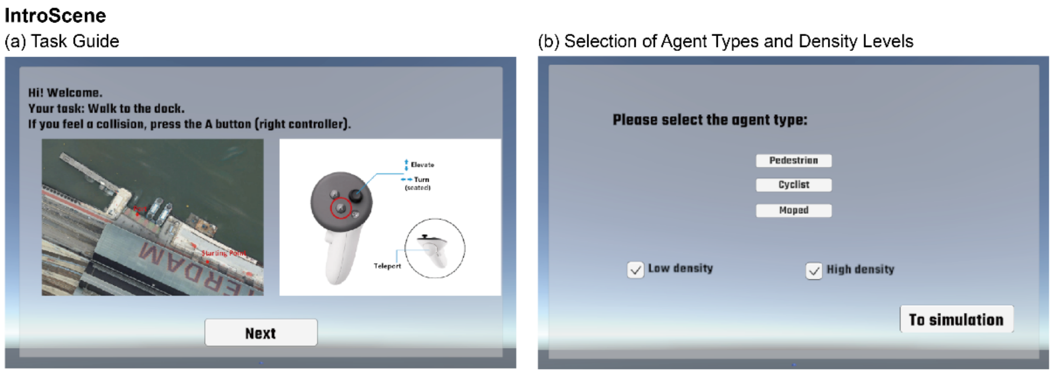

Fig 5. IntroScene. (a) Task guide; (b) Selection of agent type and density levels.

Because Linnekamp's ABM did not distinguish between different traffic density levels and only modeled the peak hour (>4000 agents were generated per hour), this study inherited his agent type distribution (58% cyclists, 35% pedestrians, and 7% mopeds) but revised the agent spawning frequency to create both high- and low-density scenarios in SimulationScene (see Fig 6). In the high-density condition, a total of 4,500 agents per hour was generated, with a spawn interval of 0.5 seconds, aligning with peak-hour simulation volumes reported in the Dutch context for urban shared spaces (Duives et al., 2019). In contrast, the low-density condition used a 2-second interval, resulting in 1,500 agents per hour (Duives et al., 2013). These configurations were designed to align the agent flow in the VR environment with the intended density levels, while maintaining system performance and computational stability. The shorter interval in the high-density scenario aimed to replicate the elevated flow rate observed in Linnekamp's peak-hour simulations. In the meantime, to maintain the runtime performance under high-density scenario, Level of Detail (LOD) was applied to avoid rendering high-resolution materials for distant objects. Character prefab textures were also compressed to minimize GPU load. These measures ensured smooth performance without significantly affecting visual clarity for behavioral analysis.

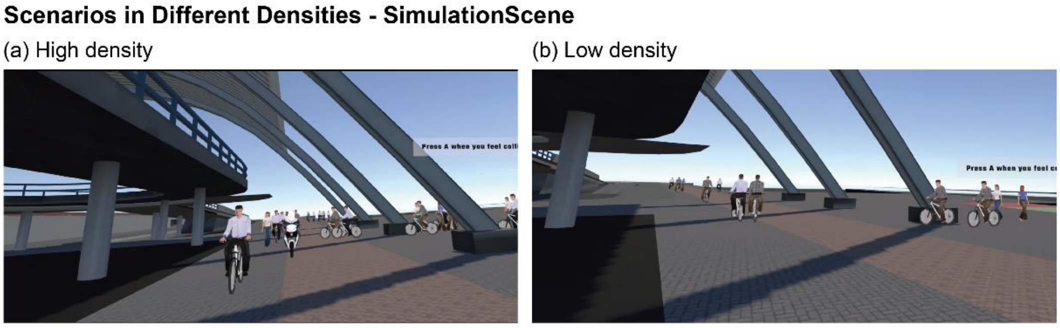

Fig 6. Scenarios in different densities - SimulationScene. (a) High density; (b) Low density.

Based on the agent type selected by each participant, SimulationScene assigned different eye heights. To match the proportions of 3D models (Unity-to-real-world scale ratio is approximately 0.94), a normal pedestrian in reality is assumed to have a height of 1.70 m and an eye height of 1.60 m, while cyclists and moped riders are assumed to have a height of 1.40 m and an eye height of 1.10 m (Landis et al., 2004). In this study, these values were adjusted in Unity so that pedestrians were set to 1.6f with eye height of 1.4f, and cyclists/moped riders were set to 1.3f with eye height of 1.1f. This means that participants choosing the pedestrian role saw other pedestrians at eye level and looked slightly down on cyclists and moped riders, while participants choosing the cyclist/moped role looked slightly up at pedestrian agents (see Fig 7).

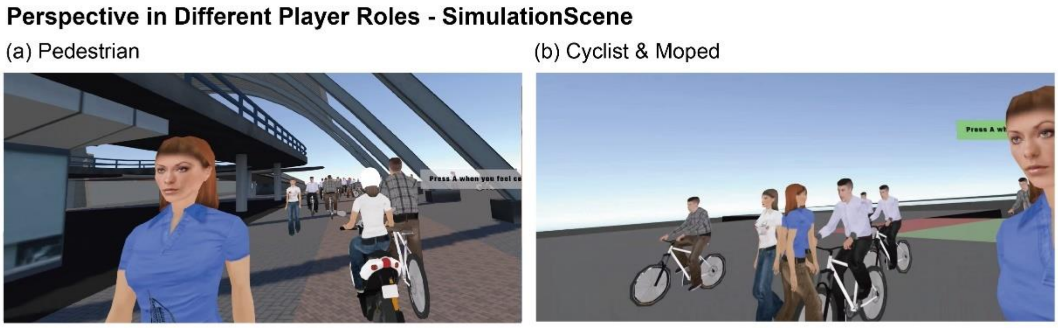

Fig 7. Perspective in different player roles – SimulationScene. (a) Pedestrian; (b) Cyclist & Moped.

Agents entered the shared space from four directions: the city side, the port side, and both directions of the bidirectional lane in front of the station. Meanwhile, the player was always spawned at a random starting point next to this lane near the Central Station and was expected to move across the flow of agents toward the ending point near the port, to simplify the simulation process (see Fig 8).

Sapwn Locations & Entering Directions

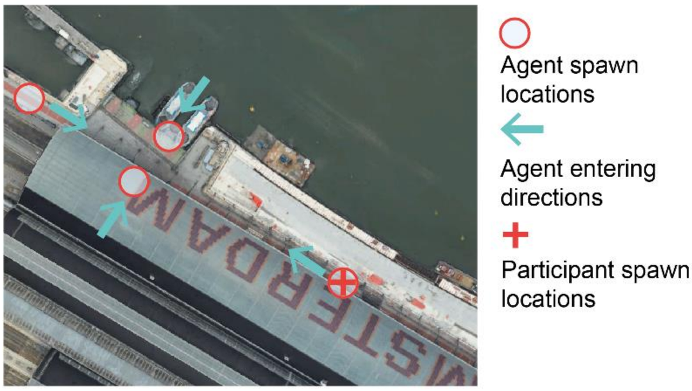

Fig 8. Agent Spawn Locations

2.2.2 ABM Replication and VR Integration

2.2.2.1 Agent and Player Attributes

In this study, each agent was defined by a set of attributes that jointly determined how it moved through space and interacted with others, including desired speed (V), radius (R), field of view (FOV). The desired speed varied by agent type (2.5-7.5 km/h for pedestrians and 5-22 km/h for cyclists/mopeds). Each agent had a fixed radius by type (0.5 m for pedestrians, 0.8 m for cyclists, and 1.0 m for mopeds), which defined its occupied space and was used for distance overlap detection. In addition, a FOV of 120 degrees enabled agents to perceive other entities and anticipate potential conflicts.

This study used agents' 3D models size in Unity as the actual sizes instead of circular diameters. For the player, randomized radii were assigned for the player, ranging from 0.8-1.0 m (pedestrians) and 1.0-1.1 m (cyclists/mopeds), to reflect variability and fit the cube-based setting of Oculus XR Integration. Considering the peripersonal space, which refers to a person's perceptional area of space that they have not yet touched but are about to touch, two types of radii were calculated in this study (see Fig 9):
(1) Physical radius: the exact size of the 3D model;
(2) Visual radius: approximately 30 cm beyond the body surface, representing the average extent of human peripersonal space reported in previous studies (Serino et al., 2015, 2018).

The introduction of a visual radius provides an anticipatory detection zone, enabling the model to represent early, perception-based reactions in close-proximity encounters. This distinction between physical and visual boundaries serves as the foundation for calculating two types of conflict probabilities in the interaction mechanism as well as the refinements in Linnekamp's ABM. Except for the distance, this study did not change the setting of other metrics to maintain simplism.

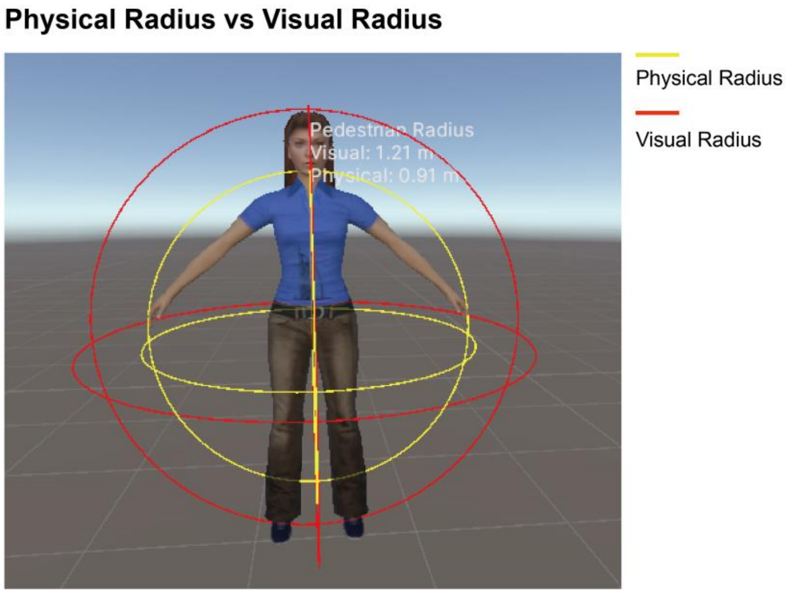

Fig 9. Physical Radius vs Visual Radius: an example of a pedestrian model

To model agent movement and interactions, the study utilized Unity's AI Navigation package (version 1.1.5). This package, built on the NavMesh system, provides real-time pathfinding and obstacle avoidance, allowing agents to follow the shortest routes and avoid environmental constraints such as pillars or other moving agents. The player's speed was controlled through Oculus XR Integration's *OVRPlayerController* script. Based on the selected type, a corresponding minimum and maximum speed as Linnekamp's ABM was assigned. These values were converted to meters per second for internal calculations in Unity. The actual player speed during movement, namely $V_{currentSpeed}$, was updated using a continuous acceleration model (see Formula 1). When the player applied joystick, the speed increased each frame based on the formula, where $accelerationRate$ was set to 2.0 m/s², $inputMagnitude$ represented the input intensity (ranging from 0 to 1), and $deltaTime$ accounted for frame time to ensure time-consistent updates:

$$V_{currentSpeed} += accelerationRate * inputMagnitude * deltaTime \tag{1}$$

2.2.2.2 Conflict Probability Modeling and Operations

As the core outcome of the interaction mechanism for assessing safety risks in shared space, the conflict probability between the player and nearby agents in this VR-integrated ABM was computed in four main steps:

Step 1 Edge Distance ($\mathbf{D_{edge}}$). The edge distance represents the shortest distance between the boundaries of the player and an agent, calculated as:

$$\mathrm{D_{edge}} = \mathrm{D_{center}} - (\mathrm{R_{player}} + \mathrm{R_{agent}}) \tag{2}$$

where $\mathrm{D_{center}}$ is the center-to-center Euclidean distance, and $\mathrm{R_{player}}$ and $\mathrm{R_{agent}}$ are their radii. A zero or negative $\mathrm{D_{edge}}$ indicates physical overlap (potential conflict).
This step is to identify whether the player and the closest agent are within potential conflict range. If $\mathrm{D_{edge}}$ is larger than 0, then conflict probability ($\mathrm{P_{conflict}}$) is set to zero, and no further computation is performed.
Step 2 Heading Difference ($\mathbf{H}$). The heading difference is the angular disparity between the movement directions of the player and the agent:

$$\Delta H = \left|\mathrm{H_{player}} + \mathrm{H_{agent}}\right|, \Delta H \in [0°, 180°] \tag{3}$$

This step is to quantify the angular disparity between movement directions. A heading difference of 90° represents the maximum crossing potential.

Step 3 Velocity Normalization. This step is to incorporate individual behavioral attributes and relative motion into conflict estimation, involving attributes related to velocities:

- Reaction time ($\mathrm{T_{reaction}}$): a random value (0–1) representing an agent's sensitivity to others.
- Relaxation time ($\mathrm{T_{relaxation}}$): a random value between 0.1 and 1 indicating how quickly the agent returns to normal speed; calculated as:

$$\mathrm{T_{relaxation}} = 1 - \mathrm{N_{FOV}}/10 \tag{4}$$

  where $\mathrm{N_{FOV}}$ is the number of agents in the field of view.
- Actual velocity ($\mathrm{V_{actual}}$): the actual velocity of an agent at current frame.

$$\mathrm{V_{actual}} = \mathrm{V_{desired}} * Clamp01(\mathrm{T_{relaxation}} - \mathrm{T_{reaction}}) \tag{5}$$

- Combined velocity ($\mathrm{V_{combined}}$): the average of the player's and agent's normalized actual speeds, representing their relative approach speed.

$$\mathrm{V_{combined}} = ((\mathrm{V_{actual,player}}/\mathrm{V_{max,player}}) + (\mathrm{V_{actual,agent}}/\mathrm{V_{max,agent}}))/2 \tag{6}$$

If $\mathrm{V_{combined}}$ is below the velocity threshold ($\mathrm{C_t}$ = 0.5), conflict probability ($\mathrm{P_{conflict}}$) is set to zero.

Step 4 Conflict Probability Computation. This step calculates the final probability combining spatial and kinematic attributes:

$$\mathrm{P_{velocity}} = Random(0, \mathrm{V_{combined}}) \tag{7}$$

$$\mathrm{P_{conflict}} = \mathrm{C_w} * \mathrm{P_{heading}} + (1 - \mathrm{C_w}) * \mathrm{P_{velocity}} \tag{8}$$

where $\mathrm{C_w}$ is the conflict weight (0.5 in this study), $\mathrm{P_{heading}}$ is derived from the heading difference, and $\mathrm{P_{velocity}}$ is proportional to $\mathrm{V_{combined}}$.

Unlike Linnekamp's original ABM, which used a fixed interaction radius, this study computed two types of conflict probabilities in real time to examine whether including peripersonal space would influence outcomes:
(1) Physical conflict probability ($\mathrm{ph\_conflictProb}$): calculated the edge distance based on 3D models' physical radii of both the closest agent and the player;
(2) Visual conflict probability ($\mathrm{vs\_conflictProb}$): calculated the edge distance by the visual radii (physical radii + 30cm) to represent the perceptual buffer of peripersonal space.
The descriptions of metrics involved calculating conflict probability are shown in Table 1.

Table 1. Agent attributes Comparison

	Linnekamp's ABM			**VR-integrated ABM**			
	Desired Speed (km/h)	**Visual Field (degree)**	**Radius (m)**	**Speed (km/h)**	**Visual Field (degree)**	**Physical Radius (m)**	**Visual Radius (m)**
Pedestrian	2.5-7.5	120	0.5	2.5-7.5	120	3D model size	3D model size+0.3
Cyclist	5-22	120	0.8	5-22	120	3D model size	3D model size+0.3
Moped	5-22	120	1.0	5-22	120	3D model size	3D model size+0.3

During the VR simulation, participants were able to report experienced conflicts in two distinct ways by UI panels (see Fig 10). First, the system calculated the physical or visual conflict probability between the player and the nearest agent in real-time. Whenever the physical conflict probability exceeded 0.75, the system triggered a UI panel that prompted the player to confirm whether they had perceived a conflict. If the player selected "Yes," the conflict was recorded as a physically triggered conflict. Second, participants recorded their perceived conflicts manually throughout the simulation by pressing the "A" button on the Oculus controller at any moment. This allowed participants to report visually perceived conflicts that were not captured by the physical model. To support the user experiences, the simulation was built on the Oculus Integration SDK.

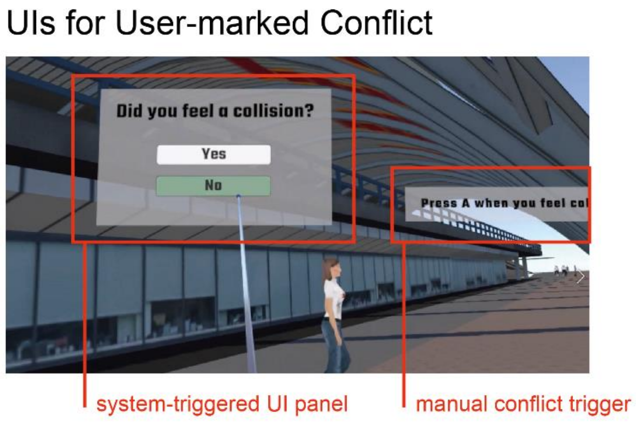

Fig 10. Two-kind UIs for collecting user-marked conflict

By calculating both physical conflict probability (based solely on model sizes) and visual conflict probability (including a perceptual buffer), and comparing each against user-marked conflicts, this study was able to test whether adding peripersonal space improves alignment between modeled and perceived conflicts, which is beneficial for determining whether Linnekamp's ABM should incorporate perceptual buffers to better reflect human anticipatory reactions in shared spaces.

2.2.2.3 Behavioral and System Performance Data Logging

Apart from the ABM building and VR integration, all user-related data were continuously collected in the background through a single core script, *MovementPattern.cs*, which served as the central logging mechanism during the experiment. This script was responsible for exporting data at regular one-second intervals into structured .csv files.

Each of the three output files corresponded to a specific aspect of the experiment: user behavior, user interaction, and system performance (see Appendix 2):

(1) MovementData.csv records player timestamp, position (x, y, z), real-time speed (converted to km/h), physical/visual conflict probability, edge distances to the nearest agent, and gap stop, a hesitation-related metric defined by periods when speed falls below 0.2 km/h (≈ 0.055 m/s). This value is consistent with thresholds used in previous pedestrian behavior studies, where speeds below 0.1–0.2 m/s are commonly regarded as an indicator of a stationary state (Kretz et al., 2006);
(2) PerformanceData.csv logs frame rate (FPS) and latency (ms) every second to monitor system performance;
(3) CollisionData.csv stores confirmed or manual conflict interactions (either press "Yes" or "A"), including time, position, agent type, conflict probabilities, and player response.

At the end of each task, when the player approached the designated ending point (within 15 meters of the dock), an ending panel would automatically appear to signal simulation completion. This served both as visual confirmation and a behavioral cue. Upon triggering the exit button, the system flushed all buffered data into local CSV files.

2.2.3 Post-Simulation Visualization Dashboard

To visualize the simulation results, this study developed a Streamlit-based dashboard. The frontend supports visualizations of logged data including system performance and movement visualizations. The dashboard read raw CSV logs (*MovementData.csv*, *CollisionData.csv*, and *PerformanceData.csv*) generated during simulation and processed these files using *pandas* and visualized using *matplotlib*, *seaborn*, and *contextily* for basemap overlays.

In performance module, plots of frame rate (FPS) and latency over time were displayed to show the VR framework's performance usability, with color-coded zones indicating safe/caution/danger ranges (see Fig 11). Latency refers to the delay between a user's action and the corresponding system response (Murray, 2013; Raaen & Kjellmo, 2015), and it is often inversely associated with FPS, as higher frame rates tend to reduce perceptible delays and improve visual smoothness (Raaen & Kjellmo, 2015). However, this relationship is not consistently reported in the literature. Some studies suggest a strong correlation between FPS and latency, while others highlight non-linear or hardware-dependent interactions (Geris et al., 2024; Janzen & Teather, 2014). This inconsistency underscores the need to treat FPS and latency as complementary but independent performance metrics. Therefore, both were measured in this study to provide a more robust and comprehensive evaluation of framework performance.

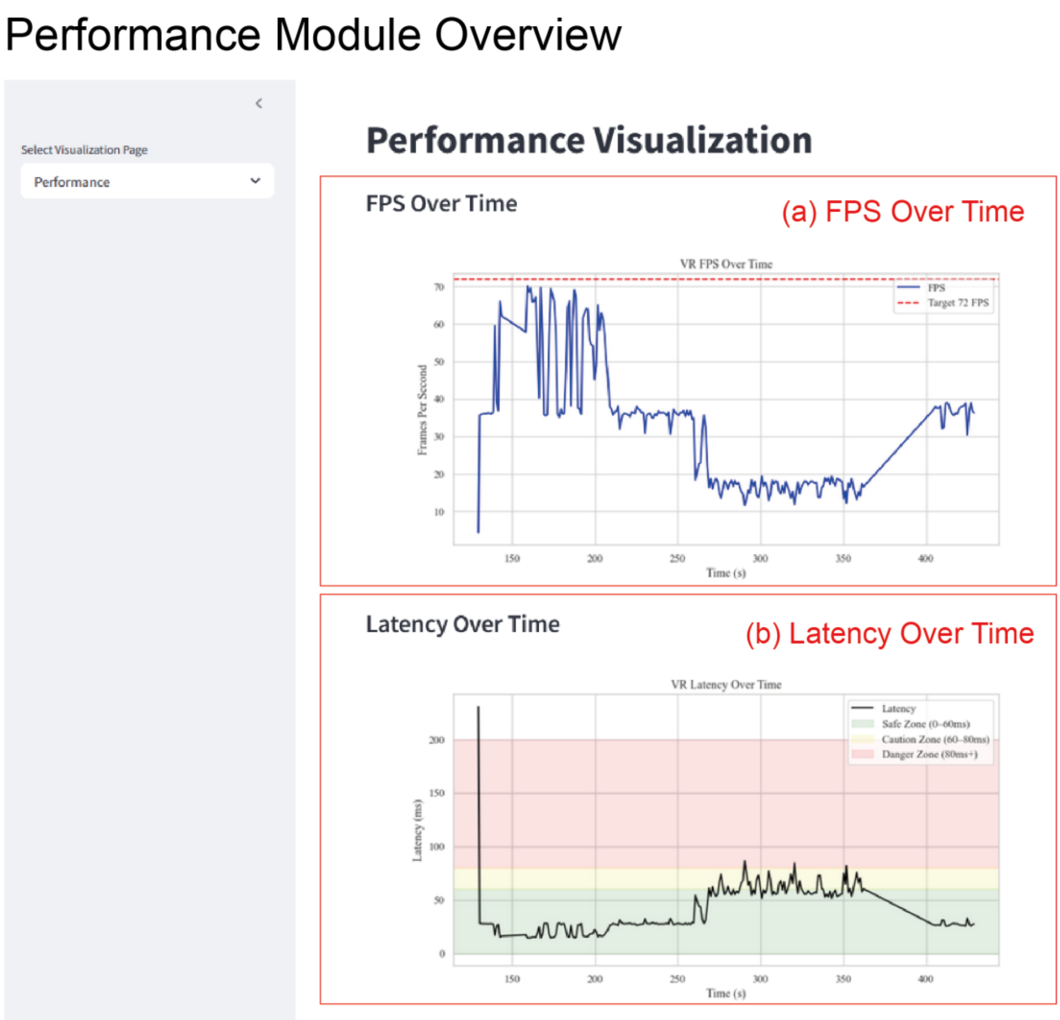

Fig 11. Performance module overview

The Movement Module integrated two types of visualizations (see Fig 12): (1) individual-level behavior plots, including movement patterns, user-marked conflicts versus edge distance, and versus player speed; (2) conflict-modeling-level plots, such as user-marked conflicts versus conflict probability and conflict probability versus gap stop. The first type provided insights into participant-agent interactions based on behavioral data, while the second type focused on validating how human perception aligned with the conflict modeling mechanism.

Movement Module Overview

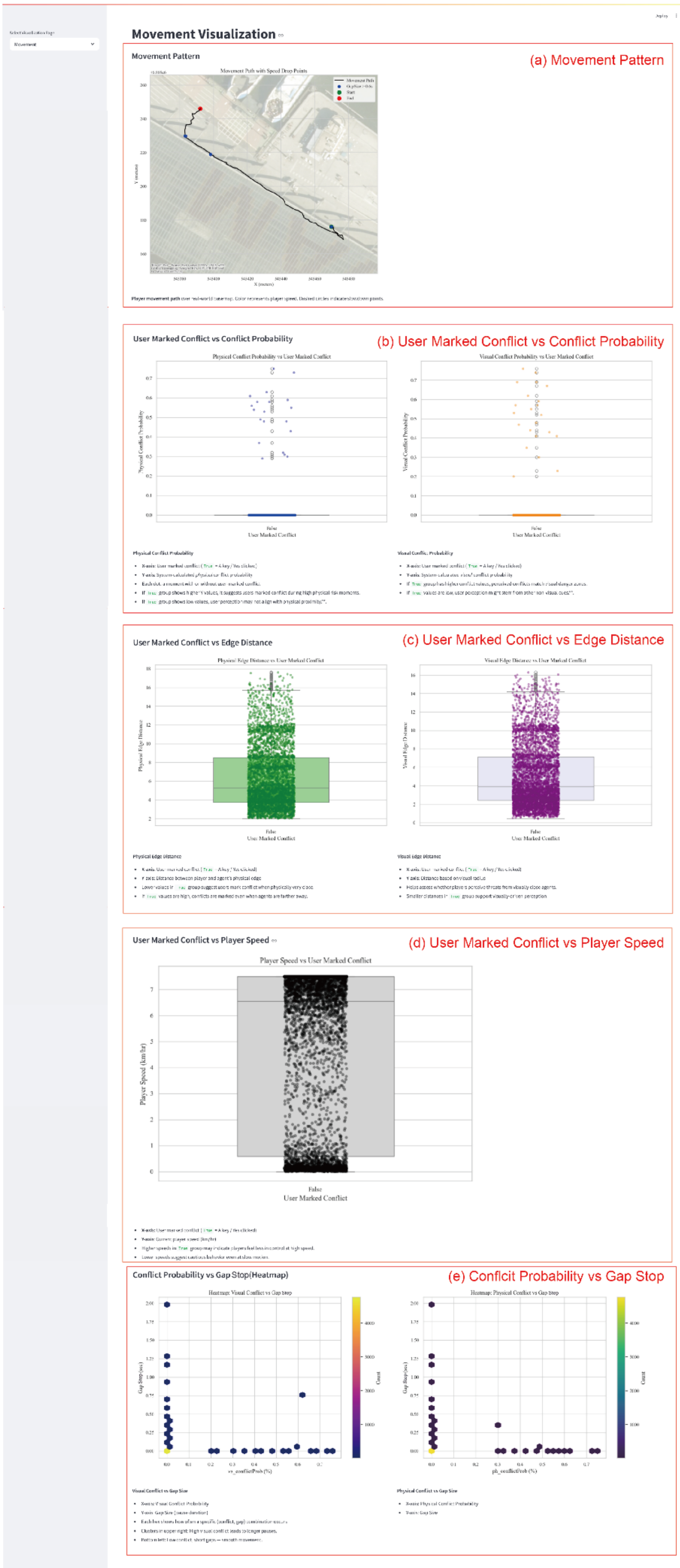

Fig 12. Movement module overview

The movement pattern plot showed the player's trajectory overlaid on a real-world basemap. Gap stop points, which were highlighted in blue dots, serving as a complementary measure to speed deceleration. In this study, gap stops larger than 0.5 seconds were considered meaningful pause events, following empirical thresholds linked to reaction or hesitation (Green, 2000; Rasouli et al., 2017). By plotting this, researchers were able to have a spatial overview of participants' movement, and places where people had hesitation. This helped with exploring if the environment setting could influence people's behavior change.

Plots relating user-marked conflicts to player speed and edge distance were used to explore how perceived conflict correlates with Linnekamp's ABM assumptions. For instance, in the speed plot, lower values in the "True" group suggest that participants were moving more slowly at the moment they perceived a conflict, potentially indicating caution or early hesitation. In contrast, in the edge distance plot, lower values for the "True" group imply that participants marked a conflict when other agents were physically closer, reinforcing the role of spatial proximity in triggering perceived risk. Understanding these patterns provides researchers with an overview of how participants behaviorally respond to perceived conflicts.

After comparing the individual-level behaviors, it would be meaningful to look at how human perceived conflicts differed with the conflict algorithm. In the plots of user-marked conflicts with physical and visual conflict probability, if the "True" group has higher visual conflict probability than physical, it suggests users are more sensitive to what they see than to actual collisions. By comparing such results, it could provide insights towards Linnekamp's conflict modelling algorithm.

As discussed earlier, conflict probability is a composite metric that integrates edge distance, speed, and heading direction. Because edge distance is already embedded in its calculation, a high conflict probability inherently reflects close physical proximity. In contrast, gap stop, pauses in movement, are not included in the computation, and therefore represents an independent behavioral indicator. This means that while edge distance and gap stop can be analyzed separately, doing so would partially duplicate the information already contained in conflict probability. For the purpose of refining a conflict-oriented ABM, it is therefore more meaningful to examine how gap stops occur at different levels of conflict probability, as this can reveal whether high-risk situations, as modeled by the ABM, actually trigger hesitation or pauses in human behavior.

To interpret this relation, the hexbin heatmap was plotted, with color intensity indicating the frequency of occurrences. Points concentrated in the lower-left suggest low perceived conflict and minimal hesitation, while distributions toward the upper-right may indicate increased pause duration in response to higher visual conflict.

2.3 QoE Evaluation

2.3.1 Participants

A total of 10 participants (60% women, 40% men) took part in the usability evaluation of the VR-integrated framework. Among all the participants (see Fig 13), most of them (90%) held a master's degree, while one participant (10%) was pursuing a PhD. Regarding age distribution, most fell into the 21-24 (30%) and 25-26 (30%) age ranges. In terms of prior experience with VR, familiarity varied: 30% had tried VR once or twice, another 30% reported occasional use, and 20% used VR regularly. Two participants (20%) had no prior experience with VR. Due to the preliminary nature of this study and time constraints, the sample size was relatively small. However, this aligns with the 'Rule of Five' of early-stage prototyping by Nielsen (2000), stating critical usability and behavioral issues can often be uncovered with as few as five users.

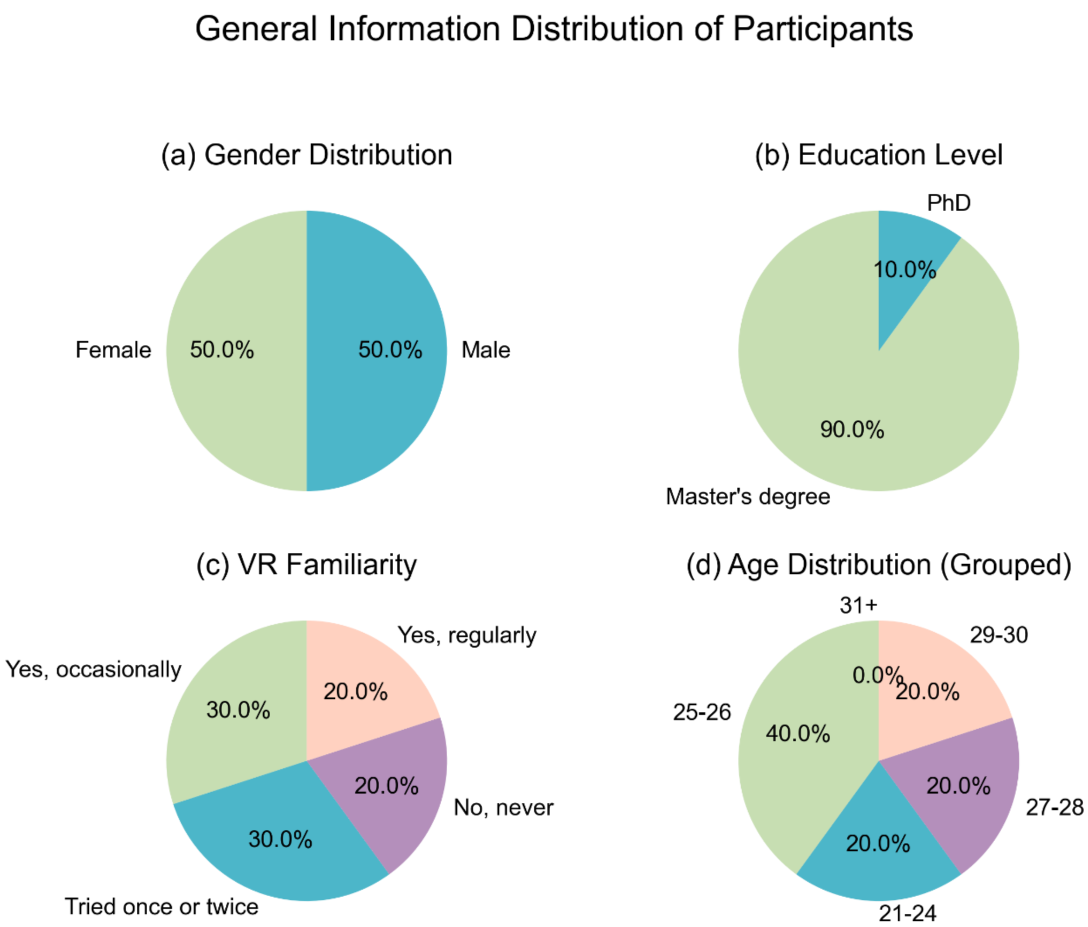

Fig 13. General Information Distribution of Participants

2.3.2 Materials

The experiment ran on a Meta Quest 3 headset connected to a personal laptop. The VR environment consisted of a training scenario with simple moving obstacles and a main simulation inspired by a real-world shared space near Amsterdam Central Station. Participants interacted using the Oculus joystick controller, which allowed manual input for perceived conflict events in addition to system-triggered prompts.

2.3.3 Procedure

The evaluation was conducted after a complete demonstration of all operations and system functionalities. The participant experience was divided into four phases: (1) a training session to familiarize users with VR operations, (2) a simulation session in a shared-space environment with behavioral data collection, (3) a demonstration of recorded movement data, and (4) a post-simulation questionnaire.

The training session (see Fig 14) introduced participants to basic VR controls in a simplified environment with a start and end point. Moving cubes approached from both sides, and participants were instructed to walk while avoiding them. When nearing a cube, a system-triggered panel prompted users to indicate if they had felt collisions. Participants could also press Button “A” to manually mark perceived, but system-unrecognized, conflicts. Upon reaching the endpoint, a confirmation panel marked the completion of training and transition to the simulation phase. A video of the training session can be found in Appendix 3.

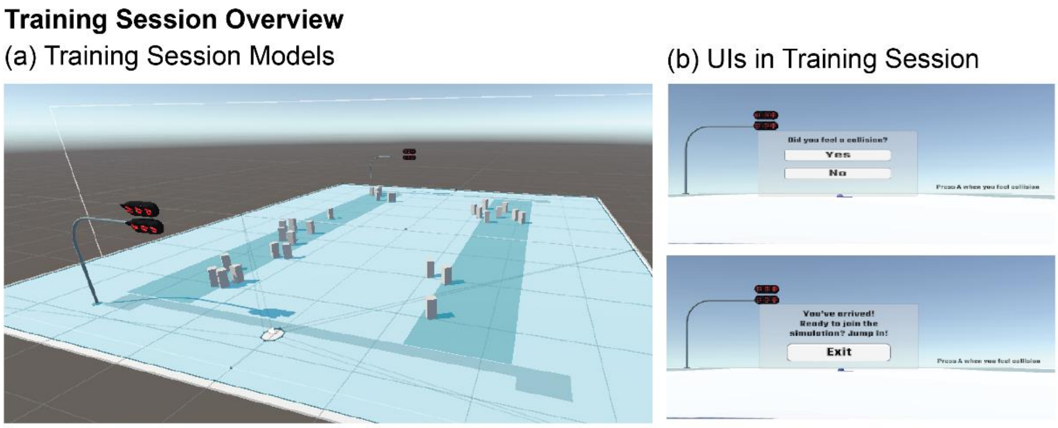

Fig 14. Training session overview. (a) Training session models; (b) UIs in training session.

Once the training session had been done, participants were transferred to the simulation session, where they were required to finish a navigation task from the spawn point to the ferry dock and their behavioral data as well as system performance data were recorded all the way. A video of the simulation session can be found in Appendix 4.

2.3.4 Measures

To evaluate QoE in the VR-based ABM framework, this study measured (1) subjective metrics of user experience (e.g., presence, usability) and (2) objective metrics of system performance (e.g., FPS, latency). Subjective data were collected via a post-use questionnaire comprising user experience items, open-ended questions, and general user information (see Appendix 5). Objective performance data were logged in Unity during the experiments.

For the user experience items, instead of administering full versions of existing questionnaires, this study selected items from multiple validated instruments to capture targeted aspects of user experience relevant to the VR task. This modular selection approach has been adopted in prior research to reduce respondent burden (Rolstad et al., 2011). Meanwhile, according to Witmer & Singer (1998), shortened or adapted scales remain valid when theoretically grounded. Four key QoE dimensions were measured using a 10-point Likert scale (1 = strongly disagree, 10 = strongly agree) (Y. M. Kim & Rhiu, 2024; Tcha-Tokey et al., 2016; Witmer & Singer, 1998; Zhao et al., 2020, 2022):

- Presence, adapted from the MEC Spatial Presence Questionnaire (MEC-SPQ) and the Presence Questionnaire (PQ), refers to the extent to which users feel physically and psychologically "present" in the virtual environment. Example items include:
  "I felt like I was really inside the virtual space."
  "The environment seemed real to me."
- Usability, derived from the System Usability Scale (SUS) and the Unified Theory of Acceptance and Use of Technology (UTAUT), measures ease of use, intuitiveness, and perceived usefulness. Example items include:
  "I found the system easy to navigate."
- Emotion and satisfaction, measured with items from the Affect Grid Questionnaire (AEQ) and Flow4D16, evaluates how enjoyable and engaging the experience was. Example items include:
  "I enjoyed using the simulation."
  "Time passed quickly while I was using the system."
- Cybersickness, assessed using negatively worded items from the Simulator Sickness Questionnaire (SSQ), captures symptoms such as nausea, dizziness, and discomfort

during VR use. Example items include:
"I felt nauseated while using the simulation."
"I experienced dizziness or eyestrain."

To minimize response bias, the items included a mix of positively and negatively worded statements (X. Zhang et al., 2016). For example, the statement "I would imagine that most people would learn to use this system very quickly" was paired with the reverse-coded item "I needed to learn a lot of things before I could get going with this system." However, for cybersickness, only negatively worded statements were used.

The open-ended questions asked participants to describe their favorite and least favorite aspects of the framework and provide impressions of the visualization dashboard. The final section collected demographic details, including age, gender, education level, and VR familiarity.

During analysis, all negatively worded items, except those for cybersickness, were reverse coded to ensure consistency in scoring. Qualitative responses were analyzed using thematic analysis following Braun and Clarke (2006). Themes were identified based on (1) Generality (mentioned by ≥2 participants), and (2) Relevance to the study's core focus (e.g., perceived usability in immersive environments), rather than frequency alone (Byrne, 2022; Vasileiou et al., 2018).

For system performance, this study logged FPS and latency in Unity during each trial and compared them between low- and high-density scenarios to examine whether environmental complexity affected system performance. Median values and variability were used due to non-normal data distributions, enabling the identification of density-related FPS differences while confirming that latency remained similar across conditions.

2.4 Generating Insights for ABM Refinement

To derive preliminary insights for refining Linnekamp's ABM, this study analyzed system and behavioral data collected from the QoE experiments. The goal was to link observed human behaviors in the VR environment to the ABM's core components - agent attributes, behavioral rules, and scenario parameters - while also evaluating the system performance of the VR-integrated ABM framework.

First, system performance (FPS and latency) was examined across traffic density scenarios to assess whether technical constraints might have influenced participant behavior. Second, individual-level behavioral patterns, including movement speed, gap stops, and movement trajectories, were analyzed to identify agent role and density differences in mobility and hesitation behaviors. Third, conflict modeling was evaluated by computing two conflict probability metrics (physical and visual) and comparing them with user-marked conflicts to assess whether system-calculated metrics align with human-perceived conflicts, and whether incorporating perceptual buffers improves this alignment. Finally, scenario-level influences were examined to understand how environmental factors, which was traffic density in this study, shaped behaviors such as personal space maintenance, speed adjustments, and perceived conflicts. This systematic analysis not only identified behavioral rules for refinement (e.g., perceptual-based conflict detection) but also highlighted how traffic densities and role types modulated behaviors, thereby providing valuable directions for improving Linnekamp's ABM's applicability.

3. Results
3.1 QoE Evaluation Results
3.1.1 User Experience Results

For user experience questions (see Fig 15), participants reported a generally high sense of presence, with overall mean scores of 7.1, suggesting that participants generally felt immersed and engaged in the virtual traffic environment. Usability-related statements had generally high mean scores among all participants (8.1). Participants agreed that most people would learn to use the system quickly (M = 7.9, SD = 1.29) and indicated a high level of clarity in repeat usage (M = 8.1, SD = 1.73). In emotion domain, most participants had positive emotional engagement. However, high standard deviations (typically > 2.5) were found across emotional items, indicating substantial individual differences. This huge difference can also be found in statements regarding cybersickness, while mean scores were generally low to moderate (e.g., nausea: M = 3.8, SD = 3.49; dizziness: M = 5.2, SD = 3.52), four participants (n = 4, 40%) explicitly reported high level of dizziness, discomfort, or motion sickness in open-ended responses. The table for all categories and descriptive statistics of the statements are listed in Appendix 6.

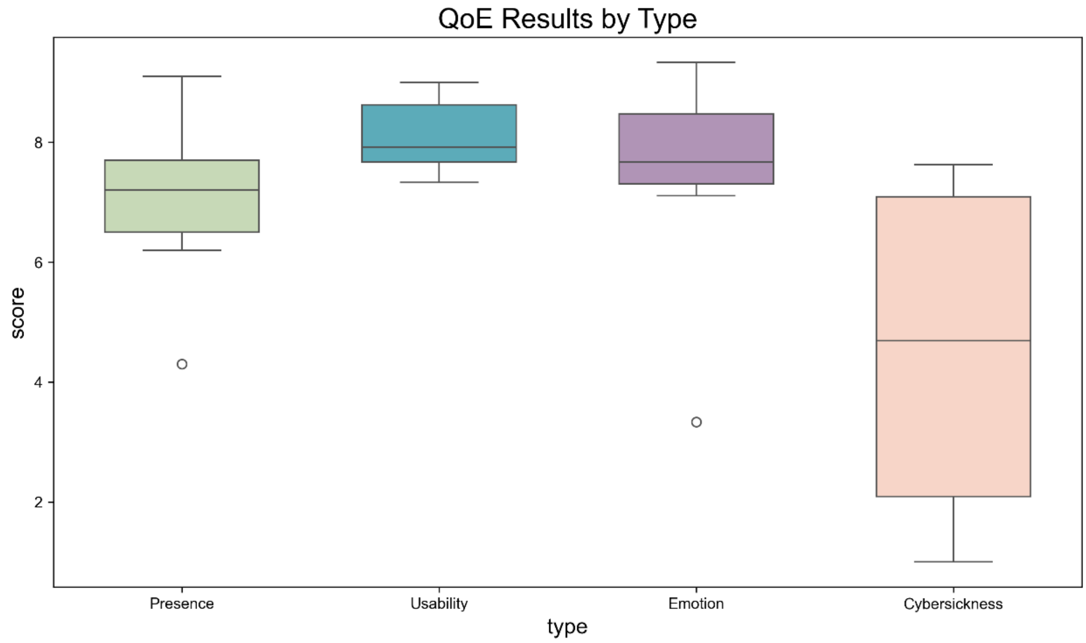

Fig 15. QoE Results by type

For open question results (see Appendix 7), positive feedback emphasized the interactive experience and clarity of visualization. Four participants (40%) highlighted interaction with the environment as their favorite feature, while three (30%) appreciated the realistic modeling of the virtual scene. The built-in tutorial was also mentioned as helpful by one participant. Regarding data visualizations, most participants (n = 6, 60%) found the outputs intuitive and easy to interpret, describing them as “explainable” and suitable for tracking transport behavior.

However, several areas for improvement were also identified. Three participants (30%) mentioned control difficulties, particularly with the hand-held controllers, describing them as less intuitive than traditional input devices. Additionally, cybersickness emerged again as a concern, with four participants (40%) reporting symptoms such as dizziness or discomfort during VR usage. This aligns with the wide variance observed in cybersickness-related questions. It is worth noting that the two participants with limited VR experience gave questions related to cybersickness relatively low scores and mentioned motion sickness in open question session.

To enhance immersion and user comfort, participants suggested multi-sensory add-ons (n = 6, 60%), including environmental sounds, vibration during collisions, and background music, to improve immersion level. Others proposed increasing environmental variation (n = 2, 20%), such as changing agent appearances and spawn behaviors. Although most users found the visualization understandable, two participants (n = 2, 20%) expressed a desire for clearer feedback during simulation (e.g., on-screen collision messages or summary metrics).

3.1.2 System Performance Results

According to the FPS distribution (see Fig 16), low-density scenarios showed a higher median FPS (Md = 58.59) compared to high-density scenarios (Md = 36.63). In the latency plot, the difference was minimal: high-density trials had a median latency of 28.81 ms, while low-density trials had a slightly higher median of 28.98 ms. Despite variability within each group, these median values suggest that frame rates decreased under higher density, while latency remained comparable across both conditions.

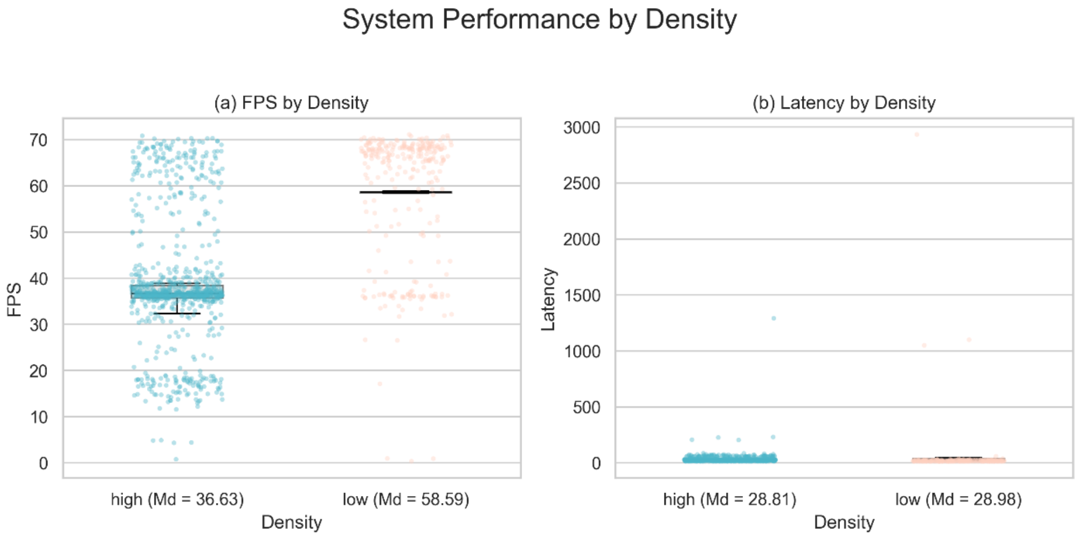

Fig 16. System performance by density

3.2 Behavioral Results

Due to the free-selection setting during the evaluation phase, participants were allowed to choose their preferred player role and scenario. As a result, the final distribution across conditions was unbalanced (see Fig 17): more participants experienced the high-density condition, and pedestrian was the most common player type. No participant selected moped. In the high-density group, five pedestrians and two cyclists participated; in the low-density group, two cyclists and one pedestrian were involved.

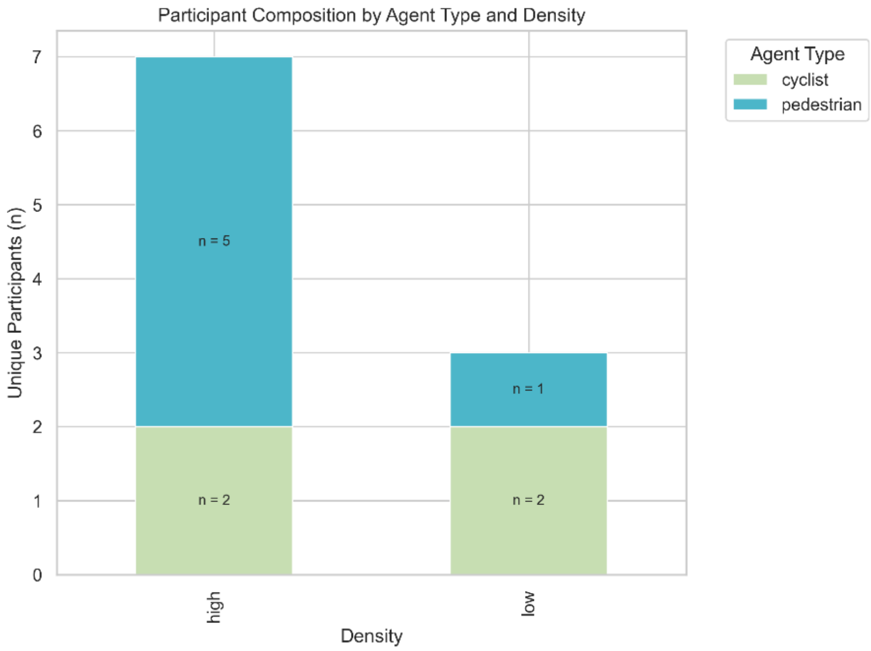

Fig 17. Participant composition by agent type and density

Considering the construction procedure of ABM, which begins with defining agent behavioral parameters and subsequently modeling interactions such as conflicts, the following analysis is organized into three domains: (1) Individual-level behavioral patterns; (2) Conflict modelling insights; and (3) Scenario-level influences. Due to the limited sample size and uneven distribution across key dimensions (e.g., density levels, player types), statistical summaries are consistently based on median values rather than means to better represent typical behaviors. In addition, for metrics with a large number of null or zero values, such as gap stop, this study filtered out zeros in order to obtain a clearer overview of the distribution and avoid distortions caused by non-informative data points.

3.3.1 Individual-level behavioral patterns

There are three parameters collected in this study that were relevant to Linnekamp's ABM agent behavioral rules: speed, gap stop, and overall movement pattern. To compare these parameters with real-human perceived conflicts, this study conducted the following analysis: (1) Movement pattern; (2) User-marked conflict vs Player speed vs Gap stop. Same as the plots in visualization dashboard, a gap stop larger than 0.5 seconds was used as the threshold for filrtering.

3.3.1.1 Movement Pattern with Gap Stop

Across participants who exhibited gap stops (>0.5s), movement patterns (see Fig 18) showed frequent stop-and-go behaviors, particularly among pedestrians in the high-density condition (e.g. P5, P10). Gap events were spatially gathered around starting point and cross road between Amsterdam Central Station and port. In high-density conditions, participants such as 6 and 10 showed dense concentrations of gap points.

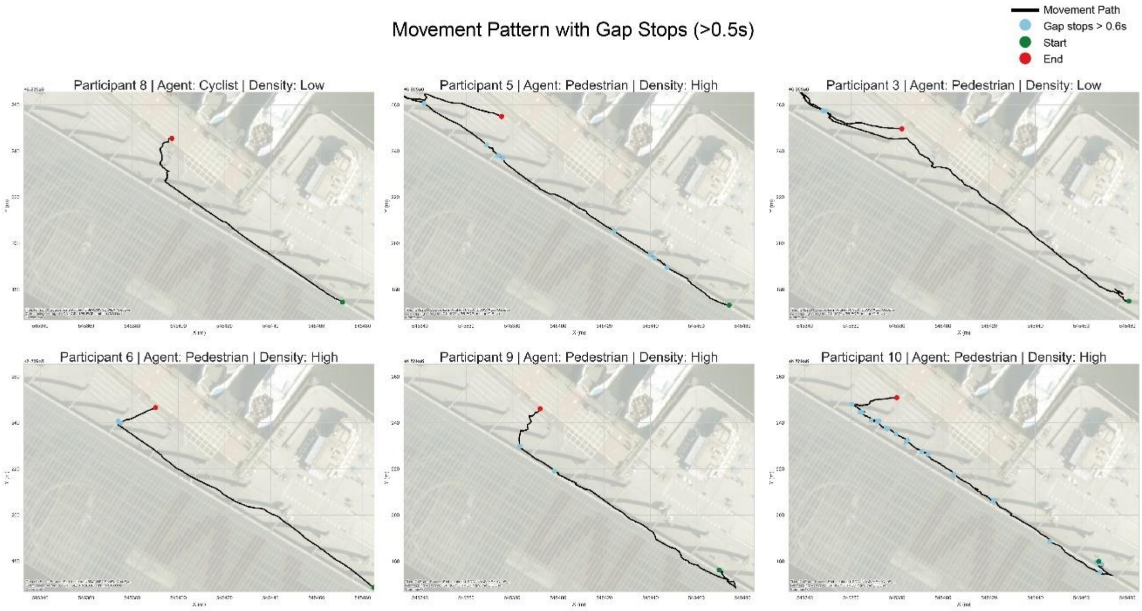

Fig 18. Movement pattern with gap stop

3.3.1.2 Perceived Conflicts with Speed and Gap Stops

This analysis investigates the relationship between user-marked conflict (shown as pink dots), player speed (gray lines), and gap stop > 0.5s (blue pans) (see Fig 19 & Fig 20). Due to the small sample size and heterogeneous agent types, the focus remains on individual-level patterns rather than aggregate trends.

Participants in high-density scenarios generally exhibited more gap stops than those in low-density scenarios, suggesting density may influence hesitation frequency. User-marked conflicts are not always associated with speed. While some participants (e.g., P2, P3, and P10) reported conflicts during low-speed or stop phases, others (e.g., P1 and P6) marked conflict even while moving at relatively high, stable speeds. Gap occurrences sometimes overlap with conflict markings, but the alignment is inconsistent. In cases like P5 and P10, certain perceived conflicts happened during gaps, suggesting some users might choose to stop in response to conflict. However, many other conflict points occurred outside gap intervals, showing that stopping is neither a universal strategy for perceived conflicts.

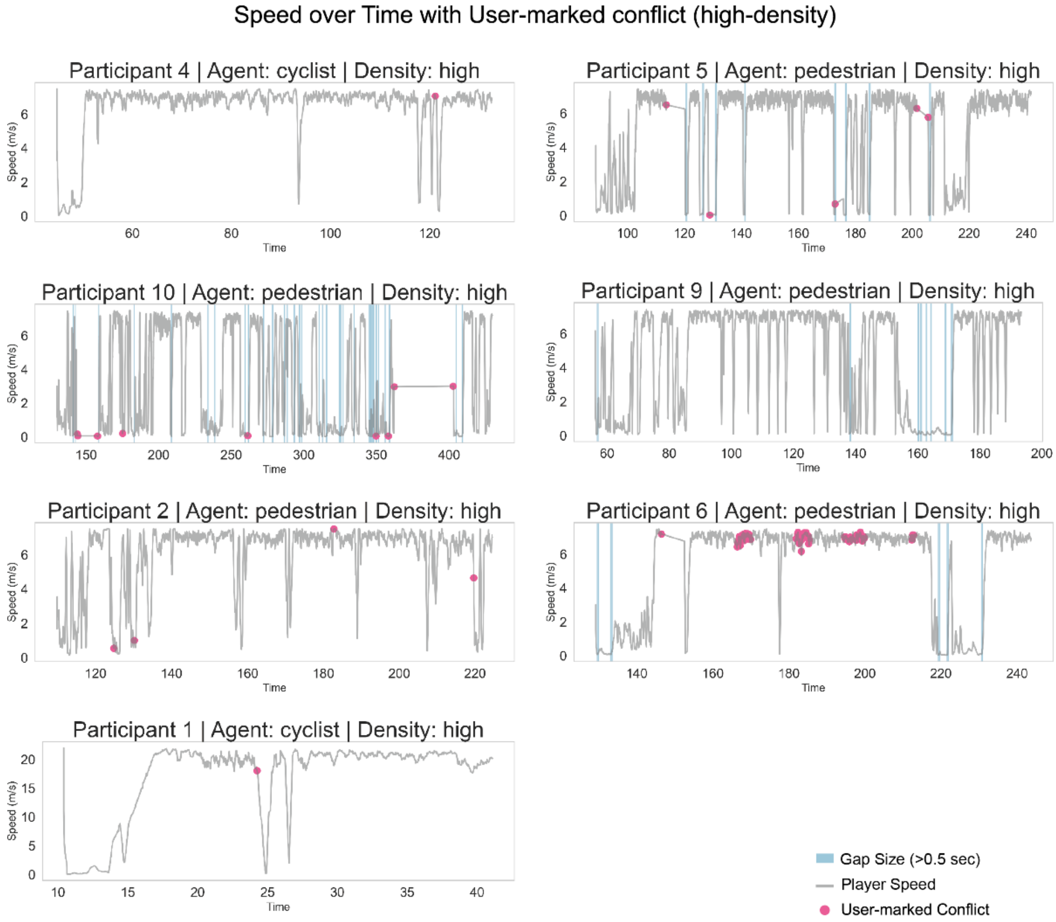

Fig 19. Speed over time with user-marked conflict (high-density)

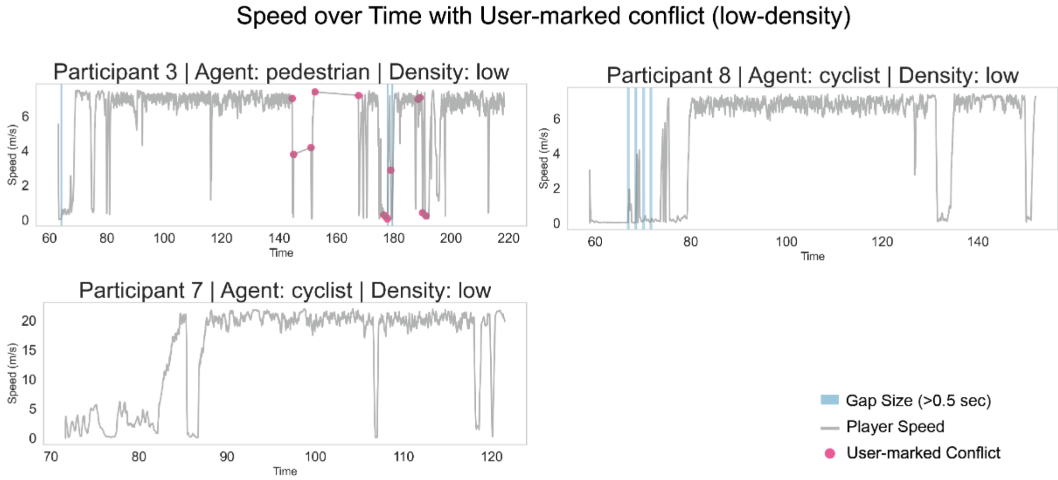

Fig 20. Speed over time with user-marked conflict (low-density)

3.3.2 Conflict modelling insights

3.3.2.1 Conflict Probability vs User-marked Conflict

In Linnekamp's ABM, a conflict is determined based on the calculated conflict probability: when the probability between two agents exceeds a threshold of 0.75, a conflict is registered. This probability is a function of several variables, including edge distance (derived from the radii of agents), heading difference, and agent speed. This study inherrited this algorithem, but calculated two kinds of conflict probabilities: physical conflict probability which was based on the mdoel boundary radius, and the visual conflict probability which was based on the peripersonal boundary radius. The aim is to assess which formulation - physical or visual - better aligns with user-marked conflicts.

The results (see Table 2) revealed an unexpected pattern: when participants marked a conflict, the system's calculated physical conflict probabilities were lower than in situations they did not mark as conflicts (mean = 0.07 vs mean = 0.15). This means the system-calculated high conflict probabilities were not perceived as conflicts by users, and the conflicts users did perceive often have relatively low calculated conflict probabilities.

For visual conflict probability, the values appeared more consistent with user-marked conflicts (mean = 0.04 when participants reported perceived conflicts, and mean = 0 when they did not). However, when comparing the results of physical and visual conflict probabilities, this pattern may be misleading. Theoretically, visual conflict probability should be equal to or higher than physical conflict probability at any given moment, since it is calculated using a larger perceptual radius. In contrast, the results showed the opposite: visual conflict probability was not only lower in magnitude (mean = 0.04) than physical conflict probability (mean = 0.07), but also occurred less frequently. In many frames where physical conflict probability was non-zero, visual conflict probability remained zero. This discrepancy is likely due to a system implementation error in the current framework, which is discussed in detail in section 5.

Table 2. Mean physical (ph_conflictProb) and visual (vs_conflictProb) conflict probabilities (non-zero only) when conflicts were marked or not. Higher means indicate greater calculated conflict probability.

Metric	Conflict Marked	Mean
ph_conflictProb (non-zero)	Marked	0.07
	Not Marked	0.15
vs_conflictProb (non-zero)	Marked	0.04
	Not Marked	0

3.3.2.2 Conflict Probability vs Gap stop

Like the visualization dashboard, this section uses hexbin heatmap (see Fig 21) to assess the relationship between physical/visual conflict probability and gap stop > 0.5s. The color intensity indicates the frequency of pause occurrences.

From the heatmaps, most gap stops occurred when both physical conflict probability and visual conflict probability were low. Similarly to the previous analysis, physical conflict probability had higher values than visual conflict probability. But for both metrics, gap stop had a generally consistent trend. When large gap stop happened, the probability metrics were low, suggesting that the length of gap stops was probably not relevant with calculated probabilities.

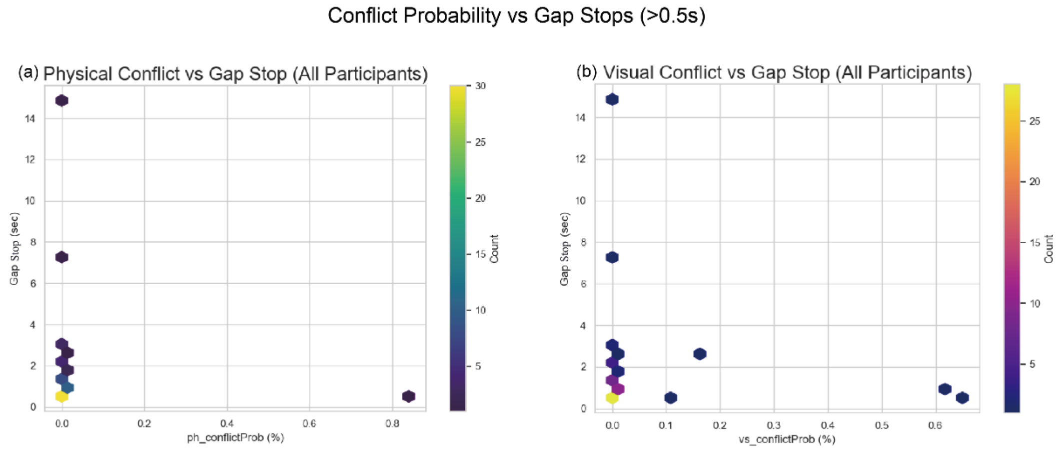

Fig 21. Conflict Probability vs Gap stop

3.3.3 Scenario-level influence

3.3.3.1 Behavioral Differences Under Density Scenarios

The comparison between high- and low-density scenarios has several differences in behavior (see Fig 22). First, participants moved slightly faster in low-density settings (Md = 7.33 km/h) compared to high-density ones (Md = 6.70 km/h). Gap stops were slightly longer in high-density conditions (Md = 0.06 s vs. 0.04 s), indicating that users are more likely to pause or hesitate when surrounded by more agents.

For conflict-relative metrics, physical conflict probability increased under high density (Md = 0.40 vs. 0.28, visual conflict probability was higher in low-density scenarios (Md = 0.47 vs. 0.38). Edge distance metrics followed a similar pattern: both physical and visual edge distances were substantially larger in low-density settings (ph: Md = 11.50 vs. 6.63; vs: Md = 10.17 vs. 5.31), indicating that users maintain broader personal space when not constrained by crowding. However, this result can also be caused by the system implementation error mentioned above, which is discussed in section 5.

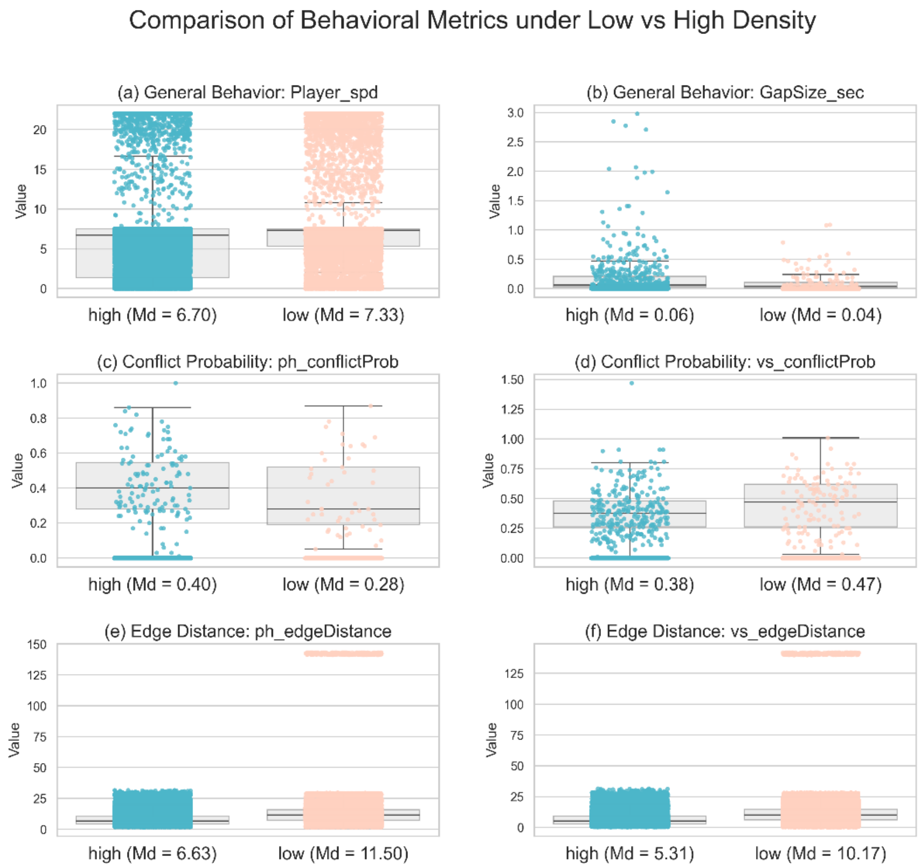

Fig 22. Comparison of Behavioral Metrics under Low vs High Density

3.3.3.2 Behavioral Differences Across Player Types

Cyclists exhibited faster and more continuous movement patterns compared to pedestrians. The median speed for cyclists was 7.5 km/h, higher than that of pedestrians at 6.73 km/h. Gap durations, which was used to capture brief pauses or deaccelerations, were shorter for cyclists (0.04 s) than for pedestrians (0.06 s), indicating fewer interruptions in movement. The median physical conflict probability was higher among cyclists (0.45) than pedestrians (0.38), while the median visual conflict probability showed the opposite trend: cyclists had a lower value (0.34) compared to pedestrians (0.39). Similarly, for participants playing the type of cyclist, visual conflict probability was lower than physical conflict probability, as found in the previous section. While this phenomenon turned opposite in the case of pedestrians.

Table 3. Statistical Description of Behavioral Differences Across Player Types

Player	**Median Speed**	**Median GapSize** (non-zero)	**Median ph_conflictProb** (non-zero)	**Median vs_conflictProb** (non-zero)
cyclist	7.5	0.04	0.45	0.34
pedestrian	6.73	0.06	0.38	0.39

3.3.3.3 Nearest Agent Type at User-Marked Conflict Moments

This section demonstrates the nearest agent type at moments when users marked a perceived conflict, categorized by density condition (see Fig 23). Under high-density conditions, a total of 88 user-marked conflicts were recorded, involving pedestrian agents most frequently (n = 53), followed by cyclists (n = 30) and mopeds (n = 4). In contrast, the low-density condition yielded only 13 user-marked conflicts, with a reversed pattern: 8 involved cyclists and 4 involved pedestrians. This suggests a shift in perceived conflict types between density conditions, where pedestrians dominate in crowded settings, while cyclists appear more prominently in sparser traffic.

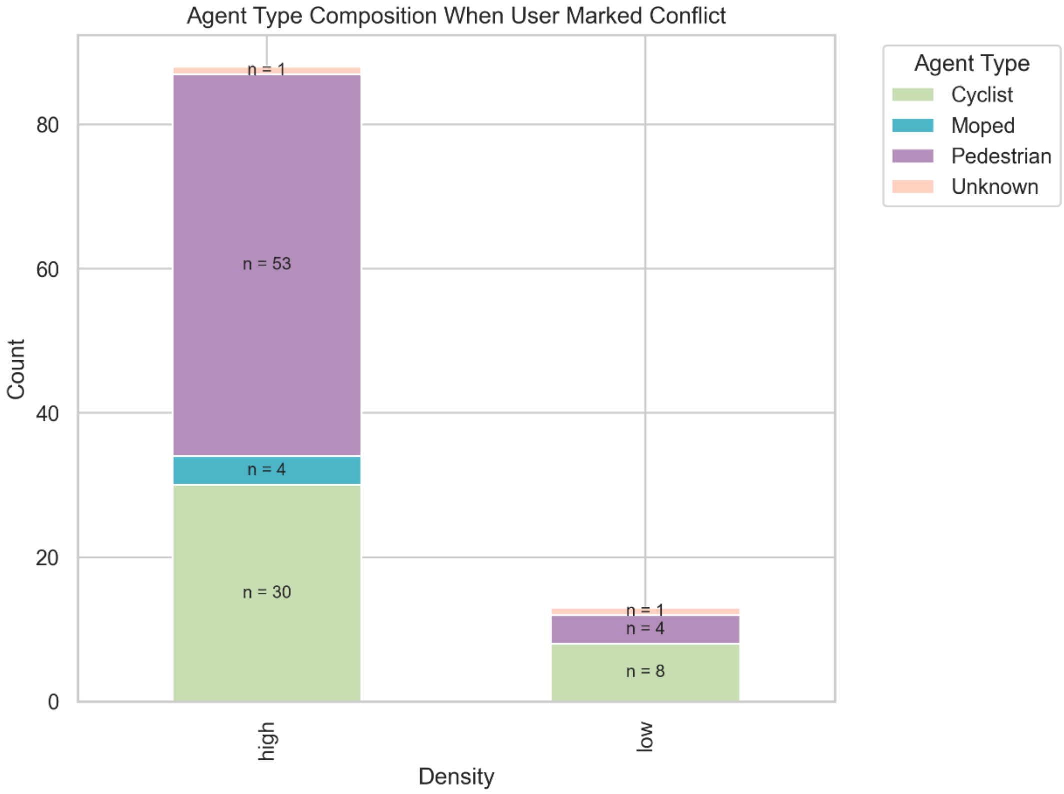

Fig 23. Agent type composition when user marked conflict

4. Discussion

This study aims to bridge the knowledge gap of how to model complex, multi-user interactions in shared spaces by integrating VR-based behavioral data collection with a researcher-facing visualization and analysis dashboard, addressing the current lack of tools for real-time experiment feedback and ABM refinement. The main research question is divided into five sub-questions: the first three focus on framework development, including integrating behavioral data collection, environmental factors, and a visualization & analysis dashboard. The fourth examines quality of experience (QoE) through a 10-participant experiment, and the fifth explores the preliminary insights to refine Linnekamp's ABM. The following sections discuss the five sub-questions respectively to demonstrate how the proposed framework addresses these methodological and practical challenges.

4.1 Development of the VR-integrated ABM Framework

4.1.1 Behavioral Logging and Conflict Detection in VR

In this VR-integrated ABM framework, behavioral data was logged in each frame, covering player position, speed (in m/s and km/h), alongside continuous calculation of visual and physical conflict probabilities based on edge distance to nearby agents by script *MovementPattern.cs*. In addition to motion data, subjective perceptions of conflict were captured in real time through in-experiment UI panels, allowing participants to actively indicate perceived conflict situations.

The technical design of the system was built using Unity and the Oculus XR Integration package, relying on OVRPlayerController for player movement. Participants navigated the scene using joystick input and rotated the view by headset, trying to mimic the real human behavior - people use their heads to look around (headset), and go forward by feet (joysticks) - to add up immersion level. All player actions were logged into structured CSV files, with movement and conflict events recorded separately, ensuring clean data separation for later analysis.

Compared to existing Unity-based behavioral logging systems, this framework offers several improvements. For instance, Feng et al. (2021) developed a VR system that recorded player waypoints, head orientation, and region transitions per frame, with data exported in JSON format. Their logging included spatial and directional information, but interaction events were only recorded upon occurrence, rather than continuously tracked. Foszner et al. (2023) also emphasized frame-by-frame trajectory recording, offering an open-source toolkit for validating tracking algorithms and supporting export in both CSV and JSON formats. Ugwitz et al. (2021) took a different approach, using Unity's trigger zones, raycasting, and keypress-based events to log user interactions. However, their default system only captured discrete interaction moments and required manual extensions to track continuous player movement. In contrast, the study combined the strengths of real-time trajectory logging with the calculation of derived behavioral metrics, such as conflict probability, and incorporated a user-facing feedback interface, supporting a more comprehensive behavioral dataset and enabling in-depth analysis of differences in behavioral and perceived conflicts across different traffic environments.

Beyond data collection, this framework was highly modularized, making the current setup can be extended to include additional environmental factors such as visibility occlusion, signage, auditory cues, or road surface conditions. Because the key data structures were defined within customizable Unity scripts, integrating new variables into the frame-by-frame logging pipeline requires less modification to the system's core architecture.

4.1.2 Density-Controlled Agent Simulation and Its Effects

This study adopted traffic density as the primary environmental factor, controlled through agent spawning frequency while keeping the agent type ratio constant. Specifically, this study followed the agent-type distribution from Linnekamp's ABM simulations (58% cyclists, 35% pedestrians, 7% mopeds) based on in-field observations, but adjusted the spawn rate to create distinct density conditions referring to literatures. In the high-density scenario, agents were spawned every 0.5 seconds (4,500 agents/hour), whereas the low-density condition used a 2-second interval (1,500 agents/hour).

The framework was implemented in Unity using the NavMesh system and AI Navigation package to control agent movement across walkable surfaces, to ensure valid pathfinding without relying on computationally intensive crowd physics models. To maintain computational efficiency, especially under high-density conditions, several graphical optimizations were applied, including the use of Level of Detail (LOD) to avoid rendering distant high-resolution materials, and compression of character textures to reduce GPU load.

Compared to existing approaches, such as the static agent number method used by Koilias et al. (2020) or the area-based density model by Dickinson et al. (2020), the dynamic, flow-based spawning strategy adopted in this study offered a more continuous and immersive representation of urban traffic. In contrast, Koilias et al.'s static quantity control is simple and suitable for small-scale simulations but lacks dynamic flow and can result in unnatural initial configurations (e.g. at the beginning of situation, there will be lots of agents crowding in the generation point). This study's approach is more aligned with Choi et al. (2024), who used Unity's NavMesh system to guide agents with varying speeds and continuous spawn rates, enabling real-time density fluctuations. However, unlike Choi et al., this study did not differentiate spawn frequency by agent type, which is a potential enhancement for future work to better reflect real-world heterogeneity in flow composition.

4.1.3 Visualization and Analysis Dashboard for Post-Experiment

To connect visualized analysis with the VR-integrated ABM framework, this study developed a Python-based visualization dashboard. Exported CSV files containing movement and interaction logs were processed in the back end and displayed through a user-friendly Streamlit web interface. This setup allowed researchers to explore participant behavior patterns, conflict frequency, and performance metrics after each experiment session. The design prioritized modularity and interpretability: data collection, processing, and visualization were handled by separate components, making it easy to adapt the system to other scenarios or input formats.

Till now, research exploring Unity and front-end integration has been still limited. In Kim et al.'s Unity-web system (2025a, 2025b), they did not directly log raw interaction data but instead route simulation outputs to pre-defined visual analytics platforms. In their work, Unity interactions were captured using UAD modes, and behavioral data was streamed to time-series databases like InfluxDB, then visualized using Grafana. These systems often incorporate point cloud representations and large language models (LLMs) to infer behavioral context and generate semantic summaries. In contrast, this study's framework focuses on post-session behavior exploration, emphasizing lightweight deployment and researcher control. While the framework does not yet integrate time-series data nor LLM analysis, its separation of modules ensures compatibility with future upgrades such as auditory stimuli, green space simulation, or LLM-assisted interpretation.

4.2 QoE of the VR-Integrated ABM Framework

The user evaluation results suggest that the VR-based ABM framework generally achieved strong usability and immersion. Most participants found the system easy to learn and engaging to use, with usability scores exceeding 8.0 and presence ratings averaging 7.1. These findings are supported by open-ended feedback, where participants highlighted the intuitive interface and clear visualization as key strengths. This matches findings from other studies. For instance, Argota Sánchez-Vaquerizo et al. (2024) found that pedestrian decisions in a VR street crossing task were highly consistent with those made in real life. This supports the idea that this VR-integrated ABM framework can be used as a substitute for real-world environments in behavior research to an extent.

However, high standard deviations in emotion- and cybersickness-related items point to substantial individual differences in user experience. While the average reported discomfort levels were moderate, 40% of participants explicitly reported symptoms such as motion sickness, particularly those with limited prior VR exposure, which is similar to results from Pöhlmann et al. (2024), who found that people with video game experience but no VR background had fewer symptoms than those with no experience at all. Beyond individual differences, system performance may also have contributed to discomfort: in high-density scenarios, the mean FPS was notably lower while latency remained relatively stable across conditions, it is a pattern consistent with VR performance studies showing that reduced rendering performance under higher system load leads to noticeable frame rate drops even when end-to-end latency does not significantly worsen (Wang et al., 2023). Such reductions in visual smoothness can aggravate motion sickness and potentially affect the accuracy and consistency of behavioral data, which is particularly relevant for this VR-integrated ABM framework that relies on participants making decisions in dynamic and complex shared space environments, even though presence and immersion scores remained acceptable.

For framework improvement, participants in this study recommended adding multisensory elements such as environmental sounds, vibration cues, and real-time feedback. Similar approaches have been shown effective in previous research. Seinfeld et al. (2022) demonstrated that combining visual, auditory, and tactile cues can reduce motion-related discomfort. Bosman et al. (2024) also found that spatial audio and ambient sound significantly enhance immersion and help maintain user attention. These insights highlight the potential of multisensory interaction to improve user experience in VR-based simulations.

On the system performance side, high traffic density scenario had lower FPS, although latency remained stable across conditions. This aligns with findings from VR performance studies, which show that reduced rendering performance (e.g., when system load increases) leads to noticeable drops in frame rates, even while end-to-end latency may not necessarily worsen significantly (Wang et al., 2023).

4.3 Insights for ABM Refinement

4.3.1 Individual behavior rules should reflect diverse and role-specific reaction strategies

In this study's result of *Perceived Conflicts with Speed and Gap Stops,* individual participants showed a wide range of behaviors when navigating shared spaces. However, Linnekamp's ABM assumes a single conflict avoidance strategy - slowing down, based on a combination of reaction time and relaxation time, which reduce speed when others appear in the agent's field of view. In contrast, in this study, some still felt conflicts even at low speeds, and not all participants slowed down when they sensed a potential conflict, suggesting that people have different reaction styles. This result aligns with the research of Ning & Li's (2022), which stated that people's collision avoidance included not only speed, but also waiting and detouring. Therefore, instead of simplifying all agents to just slow down when others appear, ABM should include a more complicated strategy module that reflects a variety of reactions, such as personal preferences.

A similar pattern was found in *Behavioral Differences Across Player Types*, where pedestrians exhibited longer gap durations than cyclists, who tended to move faster with fewer interruptions. This indicates that different roles adopt different movement strategies, supporting the idea of using type-specific perception models in ABMs. For instance, cyclists may require narrower FOV and stronger speed-related penalties, while pedestrians may respond more broadly but with delayed reactions (Mastora et al., 2023).

4.3.2 Conflict probability modeling should incorporate perceptual and contextual information

As to conflict probability modelling mechanism, this study found an interesting result: people felt more “being conflicted” when the physical conflict probability was low. To explain this, there could be two possibilities: (1) Overly strict and simplified algorithm for calculating conflict probability; (2) Human perception delay; (3) Lack of environmental information.

Back to Linnekamp’s ABM, the conflict probability was dominated by a series of strict requirements: (1) There must have been a physical overlap between two agents’ radius; (2) The combined velocity is larger than 0.5 (the threshold of Ct). If one requirement fails, the conflict probability returns to 0. Also, this algorithm mainly focused on instantaneous parameters such as distance, speed, and the number of agents per frame. While Jiao et al. (2025) pointed out that the context information, especially the accumulated number of agents in view, can also influenced the conflict probability – for instance, when people see lots of agents coming to them, it is more likely to have a conflict due to nervousness.

On the other hand, human perception is not always aligned with objective conflict – they tend to perceive the conflict before or after it physically occurs (Kaß et al., 2018), making the result seems against common sense. Participants also noted that missing environmental information (e.g., sounds of vehicles) affected their behavior and perception in the QoE questionnaire, particularly for side or rear conflicts (Payzan-LeNestour et al., 2021; Sayin et al., 2015). This can also be a cause for the wicked result of comparing user-marked conflict with conflict probabilities and gap stops. To improve ABM, it would be valuable to incorporate contextual and historical information (e.g. the speed of agents, and the number of agents in the player’s field of view over the past few frames) into the conflict modelling mechanism.

Another aspect worth discussing is the differences between perceived conflict and the conflict that indeed happens. As mentioned above, human perception might not be the perfect substitute to be accounted as “ground truth”, but it significantly shapes behavioral responses (Ihssian & Ismail, 2023). This study suggests taking this human perception data as a behavior trigger, it is more important to know what the mechanism is between human perception, their reactions based on the perception, and finally the consequences of the reactions (e.g. conflict). Further study is needed to explore the role of perceived conflict in shared space interactions (e.g. “*How does behavior change when a conflict is perceived in a complex interaction environment*?” or “*Under what conditions is conflict more easily perceived?*”), and to corporate such perception-behavior interaction into the agent rules and the calculation of conflict probability.

4.3.3 Scenario-aware rules should be added to simulate density-based behavioral variation

When comparing different density scenarios, participants showed lower speeds and longer hesitation gaps in low density scenario, suggesting a more cautious movement style. This aligns with findings from Sieben et al. (2017), who reported increased hesitation and avoidance behavior under crowded environments. Such results indicate that high-density settings naturally induce behavioral hesitation.

User-marked conflicts towards different agent types also depend on density. In high-density scenarios, perceived conflicts were often caused by nearby pedestrians, while in low-density settings, they were more often linked to fast-approaching cyclists, even though the proportion of each agent type was fixed. This pattern may be explained by different underlying reaction mechanisms. As noted in Appert-Rolland et al. (2018), in low-density environments, individuals tend to react more strongly to fast or sudden-moving objects, likely because such movements stand out more against an otherwise calm background. In contrast, under high-density conditions, people are more sensitive to the mental burden caused by immediate proximity and frequent micro-interactions, leading to increased perceived conflict by nearby agents regardless of speed (Engelniederhammer et al., 2019). Therefore, when improving ABM systems, it may be useful to implement density-sensitive behavioral rules. For instance, in low-density contexts, agents could have enhanced responsiveness to fast-moving objects by shortening reaction time or assigning higher conflict weight to high-speed agents. In high-density scenarios, emphasis could shift toward proximity-based hesitation or increased frequency of micro-conflict detection.

To summarize, there are three possible directions to improve Linnekamps' ABM. First, agent behaviors should reflect the diversity of human strategies beyond simple deceleration. Second, conflict modeling should go beyond strict geometric thresholds to include perceptual delay, visual load, and environmental cues. Further research about the mechanism between perception and reaction should be emphasized. Third, behavioral rules should be context-aware: agents in low-density settings may need faster responses to sudden movements, while high-density scenarios require greater sensitivity to proximity and micro-interactions.

5. Limitations

5.1 Development Constraints: Behavioral Data Accuracy and Interaction Timing

Several technical limitations during the development and implementation of the VR system may have influenced user behavior and data accuracy. A key technical issue was the misalignment between visual conflict probability and physical conflict probability, as shown in section 3.2.2.2 and 3.2.2.3. Naturally, visual conflict probability should be larger than the physical conflict probability at any frame. While in this study, the visual result was also occasionally lower than physical conflict probability. This can be caused by the logging sequence, where the physical conflict probability was first logged, then the visual conflict probability. Although there is a sequence, they should both be logged within the same frame. The time sequence can result in a lower visual conflict probability.

In addition, the UI panel for user feedback occasionally appeared with a delay, causing participants to miss the intended response window. This not only introduced temporal mismatches but also disrupted participants' natural behavior. For future development, it may be more effective to remove the conflict panel altogether and rely solely on user-initiated conflict markings. This would reduce timing inconsistencies and prevent interruptions caused by sudden interface pop-ups, allowing participants to behave more naturally and realistically within the simulated environment.

Last, although different agent types (pedestrian vs cyclist/moped) were assigned different eye heights in the script and verified in the Unity console, the height differences were still not visually obvious from the VR perspective. This may have influenced how participants observed the environment and how they reacted (Bailenson et al., 2001). Further implement to increase the vertical differentiation between agent types is needed.

5.2 QoE Constraints: Realism, Interaction, and Cybersickness

From the questionnaire results, the motion model of agents, based on Unity's NavMesh package, lacked smooth turning and realistic speed changes, which some participants noted as unrealistic. This limited behavioral credibility may have reduced user trust and affected their willingness to engage with the agents meaningfully. High-resolution agents and more smooth movement should be adapted in later version of this framework. Apart from this, multi-sensory modules (e.g. sounds) and more flexible way of operation (e.g. using joystick to rotate view), can be added to the framework to improve QoE.

Another frequently reported issue was cybersickness. One contributing factor was that several participants were first-time VR users; their lack of familiarity with immersive environments likely increased their susceptibility to motion sickness (da Silva Marinho et al., 2022). The other reason was relevant to the system performance. system performance played a role. Under high-density scenarios, the FPS became unstable and latency increased, both of which are known to aggravate cybersickness symptoms (C. Zhang, 2020). This can be caused by the laptop's computation capacity.

5.3 Insight Constraints: Sample Size and Condition Distribution

This study hosted a small-scale experiment for getting real participants' behavioral data for ABM refinement. However, there are two constraints that may hinder the collection of comprehensive insights: First, despite the preliminary results, it is important to acknowledge that the limited participant sample in this study may have introduced measurement errors and explanatory bias. Future research should involve a larger and more diverse sample to enhance the generalizability and robustness of the findings. Second, in the experiment design, participants were free to choose their preferred agent type and density. However, in a small-sample experiment, this setting does not easily produce a balanced distribution of conditions for ABM refinement. For example, in this study, far more participants chose the high-density condition than the low-density condition, and pedestrians were selected most frequently, while no participant chose the moped agent type. In a small-sample study, pre-assigning each participant's role and density would clearly result in a more evenly distributed dataset.

6. Future Suggestions

Future development of the VR-integrated framework can proceed in several directions: (1) Functional optimization and compatibility expansion; (2) Diversification of experimental design; (3) Expanded applications.

For function optimization and compatibility expansion, first, introducing a reaction time buffer could help mitigate the current issue of timestamp mismatches between system-recorded and user-marked behaviors. Second, the current system is limited to visual field and proximity-based cues. Integrating multisensory inputs, such as directional sound cues, peripheral blur, or partial occlusion, could better simulate real-world perception and enhance the validity of conflict detection mechanisms. In addition, the field of view and camera height could be further adjusted to create clearer distinctions between agent types, improving participants' depth perception and spatial awareness, which may in turn lead to more realistic behavioral responses. Third, for data collection, integrating eye-tracking or physiological sensors (e.g., heart rate monitors) would allow for more objective detection of user attention and stress levels during conflict episodes. Then, migrating the platform to OpenXR, an open-standard API for VR/AR development, is highly recommended. Unlike Oculus Integration, OpenXR does not provide pre-packaged scripts (e.g., OVRPlayerController), but it offers long-term support, integrates more seamlessly with Unity's toolchain, and enables cross-platform development. This would allow the framework to be deployed on various headsets such as Pico or HTC Vive, facilitating broader user testing and application in different environments. Finally, for better system performance and less cybersickness, computers with powerful computation capabilities are highly recommended.

On the experimental design side, more complex traffic interaction scenarios should be introduced. The current setup mainly involves linear movement and a limited number of agent types. Future iterations could include bidirectional flows, multiple simultaneous tasks, alternating between conflict and non-conflict missions, and various types of agents (e.g. pedestrian, moped, cyclist, driver, scooter riders) to simulate cooperative or competitive dynamics. Moreover, for future experiments, the participant pool should be diversified beyond university students. Including elderly individuals, children, or habitual cyclists would offer deeper insights into how different user groups perceive and respond to conflict in shared spaces. Additionally, facilities with higher computational capacity are generally preferred, as they improve system smoothness and reduce the risk of motion sickness.

For future expansions, calibration and validation modules can be added, where behavioral data captured from VR sessions can be used to calibrate agent-based models, moving beyond rule-based assumptions toward empirically grounded simulation. In the long run, the VR framework could be extended as a generalized tool for testing more mobility scenarios, beyond the scope of shared space, allowing users to experience different rule settings, such as right-of-way schemes or yielding behavior, and gather behavioral feedback and strategic preferences.

7. Conclusion

This study developed a VR-integrated ABM framework using Linnekamp's ABM as a case study, aiming to collect human behavioral data, provide varied environmental settings, and support immediate visualization of data analysis. Through an experiment, the study evaluated the QoE of this framework and compared the results with Linnekamp's ABM to generate insights for refinement.

In terms of framework, this study introduced several technical and methodological enhancements, including continuous behavioral logging with dual conflict probability calculations (physical and visual), in-experiment UI panels for real-time perceived conflict reporting, and density-controlled agent spawning. The modular Unity architecture allowed environmental configuration (e.g., agent ratio, spawn rate) and streamlined CSV data export. A Python-based visualization dashboard enabled immediate post-session analysis, allowing researchers to inspect behavioral patterns and performance metrics without extra data processing.

For QoE, higher traffic density reduced frame rates but had little effect on latency. The framework achieved high presence and usability, with generally positive emotional engagement, though 40%, especially people unfamiliar with VR, reported cybersickness. Feedback praised the interactive environment and visualizations but noted controller usability issues and lack of sensory feedback, suggesting multi-sensory elements and greater environmental complexity.

Behavioral data showed context-sensitive, strategy-diverse conflict responses, with greater hesitation in high-density settings. Perceptions were influenced by physical metrics, agent types, and environmental cues. Compared with Linnekamp's ABM, which assumed only one avoidance strategy (slowing down) and fixed conflict thresholds, results highlight the need for perception-driven, type-specific, and context-aware modeling.

Limitations included logging mismatches between visual and physical conflict probabilities, delayed UI responses, insufficient agent height differences, limited motion realism, lack of sensory feedback, and unstable FPS in high-density scenarios. The small, self-selecting participant sample also limited generalizability.

Future improvements should address technical stability, add multisensory cues, refine agent realism, integrate eye-tracking and physiological sensors, migrate to OpenXR, and use higher-performance hardware. Expanding traffic scenarios, agent diversity, and participant demographics would improve ecological validity. Broader applications include calibration and validation modules for ABM refinement and extending the framework to other mobility contexts.

In summary, this early-stage prototype integrates VR-based behavioral data collection, environmental variation, and post-session visualization to support empirical ABM refinement. It shows VR's potential to simulate complex traffic interactions and analyze behavioral data for model improvement, offering a foundation for more realistic and human-centered traffic simulations.

Statement of AI

Artificial intelligence (AI) assistance was used for language polishing and programming grammar checking. For language polishing, it was applied to correct grammar and improve the readability of the manuscript. For programming grammar checking, it helped identify and fix bugs in C# scripts and provided more efficient solutions. No AI tools were involved in the study design, data collection, analysis, or interpretation of results.

References

Aguilar, L., Gath-Morad, M., Grübel, J., Ermatinger, J., Zhao, H., Wehrli, S., Sumner, R. W., Zhang, C., Helbing, D., & Hölscher, C. (2024). Experiments as Code and its application to VR studies in human-building interaction. *Scientific Reports*, *14*(1), 9883.

Aguilera-García, Á., Gomez, J., Sobrino, N., & Vinagre Díaz, J. J. (2021). Moped scooter sharing: Citizens' perceptions, users' behavior, and implications for urban mobility. *Sustainability*, *13*(12), 6886.

Alexander, M. (2015). *Analyzing the behavior of cyclists at intersections to improve behavior variability within micro-simulation traffic models*.

Allocated, L., & Core, C. (2013). Streets as public spaces and drivers of urban prosperity. *Of Urban Prosperity*, 108.

Angioi, F., & Bassani, M. (2022). The implications of situation and route familiarity for driver-pedestrian interaction at uncontrolled mid-block crosswalks. *Transportation Research Part F: Traffic Psychology and Behaviour*, *90*, 287–299. https://doi.org/10.1016/j.trf.2022.09.003

Angulo, A. V., Robartes, E., Guo, X., Chen, T. D., Heydarian, A., & Smith, B. L. (2024). Evaluating current and future pedestrian mid-block crossing safety treatments using virtual reality simulation. *Accident Analysis & Prevention*, *206*, 107715.

Anvari, B., Bell, M. G., Sivakumar, A., & Ochieng, W. Y. (2015). Modelling shared space users via rule-based social force model. *Transportation Research Part C: Emerging Technologies*, *51*, 83–103.

Appert-Rolland, C., Pettré, J., Olivier, A.-H., Warren, W., Duigou-Majumdar, A., Pinsard, E., & Nicolas, A. (2018). Experimental study of collective pedestrian dynamics. *arXiv Preprint* *arXiv:1809.06817*.

Argota Sánchez-Vaquerizo, J., Hausladen, C. I., Mahajan, S., Matter, M., Siebenmann, M., van Eggermond, M. A., & Helbing, D. (2024). A virtual reality experiment to study pedestrian perception of future street scenarios. *Scientific Reports*, *14*(1), 4571.

Ayres, G., & Mehmood, R. (2009). *On discovering road traffic information using virtual reality simulations*. 411–416.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments*, *10*(6), 583–598.

Balmer, M., Meister, K., Rieser, M., Nagel, K., & Axhausen, K. W. (2008). Agent-based simulation of travel demand: Structure and computational performance of MATSim-T. *Arbeitsberichte Verkehrs-Und Raumplanung*, *504*.

Barceló, J. (2010). *Fundamentals of traffic simulation* (Vol. 145). Springer.

Basak, K., Hetu, S. N., Azevedo, C. L., Loganathan, H., Toledo, T., & Ben-Akiva, M. (2013). *Modeling reaction time within a traffic simulation model*. 302–309.

Batista, M., & Friedrich, B. (2022). Investigating spatial behaviour in different types of shared space. *Transportation Research Procedia*, *60*, 44–51.

Bazghandi, A. (2012). Techniques, advantages and problems of agent based modeling for traffic simulation. *International Journal of Computer Science Issues (IJCSI)*, *9*(1), 115.

Bella, F. (2011). How traffic conditions affect driver behavior in passing maneuver. *Advances in Transportation Studies*, *Special Issue 2011*. https://onlinepubs.trb.org/onlinepubs/conferences/2011/RSS/1/Bella,F.pdf

Bella, F., & Silvestri, M. (2015). Effects of safety measures on driver's speed behavior at pedestrian crossings. *Accident Analysis & Prevention*, *83*, 111–124.

Bhagavathula, R., Williams, B., Owens, J., & Gibbons, R. (2018). *The reality of virtual reality: A comparison of pedestrian behavior in real and virtual environments*. *62*(1), 2056–2060.

Bianchi, C., Cirillo, P., Gallegati, M., & Vagliasindi, P. A. (2007). Validating and calibrating agent-based models: A case study. *Computational Economics*, *30*, 245–264.

Blissing, B., Bruzelius, F., & Eriksson, O. (2019). Driver behavior in mixed and virtual reality–a comparative study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *61*, 229–237.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl_3), 7280–7287.

Bosman, I. de V., Buruk, O. 'Oz,' Jørgensen, K., & Hamari, J. (2024). The effect of audio on the experience in virtual reality: A scoping review. *Behaviour & Information Technology*, *43*(1), 165–199.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, *52*, 455–463.

Burghout, W., Koutsopoulos, H. N., & Andreasson, I. (2005). Hybrid mesoscopic–microscopic traffic simulation. *Transportation Research Record*, *1934*(1), 218–225.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis.

Quality & Quantity, *56*(3), 1391–1412.

CGTrader—3D Model Store. (n.d.). CGTrader. Retrieved June 22, 2025, from https://www.cgtrader.com/

Chamilothori, K., Wienold, J., & Andersen, M. (2019). Adequacy of immersive virtual reality for the perception of daylit spaces: Comparison of real and virtual environments. *Leukos*, *15*(2–3), 203–226.

Cheliotis, K. (2022). *Simulating Common Indoor Crowd Behaviours in 3D Environments Using ABMU*. 1–4.

Chen, C., Zhao, X., Liu, H., Ren, G., Zhang, Y., & Liu, X. (2019). Assessing the influence of adverse weather on traffic flow characteristics using a driving simulator and VISSIM. *Sustainability*, *11*(3), 830.

Choi, C.-H., Park, S.-W., Park, J., Kim, J.-H., Kim, J.-S., Yang, H.-S., Lee, B.-E., Jung, S.-H., & Sim, C.-B. (2024). Prediction and prevention of crowd-crush accidents using crowd-density simulation based on unity engine. *Discover Applied Sciences*, *7*(1), 33.

da Silva Marinho, A., Terton, U., & Jones, C. M. (2022). Cybersickness and postural stability of first time VR users playing VR videogames. *Applied Ergonomics*, *101*, 103698.

D'Haese, S., Van Dyck, D., De Bourdeaudhuij, I., Deforche, B., & Cardon, G. (2014). The association between objective walkability, neighborhood socio-economic status, and physical activity in Belgian children. *International Journal of Behavioral Nutrition and Physical Activity*, *11*, 1–8.

Dias, C., Iryo-Asano, M., Nishiuchi, H., & Todoroki, T. (2018). Calibrating a social force based model for simulating personal mobility vehicles and pedestrian mixed traffic. *Simulation*

Modelling Practice and Theory, *87*, 395–411. https://doi.org/10.1016/j.simpat.2018.08.002

Drut, M. (2018). Spatial issues revisited: The role of shared transportation modes. *Transport Policy*, *66*, 85–95.

Duives, D. C., Daamen, W., & Hoogendoorn, S. P. (2013). State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies*, *37*, 193–209.

Duives, D. C., Sparnaaij, M., Daamen, W., & Hoogendoorn, S. P. (2019). How many people can simultaneously move through a pedestrian space? The impact of complex flow situations on the shape of the fundamental diagram. *arXiv Preprint arXiv:1908.07208*.

Edquist, J., & Corben, B. (2012). Potential application of Shared Space principles in urban road design: Effects on safety and amenity. *Report to the NRMA-ACT Road Safety Trust, MONASH University, Accident Research Centre*.

Engelniederhammer, A., Papastefanou, G., & Xiang, L. (2019). Crowding density in urban environment and its effects on emotional responding of pedestrians: Using wearable device technology with sensors capturing proximity and psychophysiological emotion responses while walking in the street. *Journal of Human Behavior in the Social Environment*, *29*(5), 630–646.

Feng, Y., Duives, D., & Hoogendoorn, S. (2021). Development of a VR tool to study pedestrian route and exit choice behaviour in a multi-story building. *arXiv Preprint arXiv:2103.05560*.

Foszner, P., Szczęsna, A., Ciampi, L., Messina, N., Cygan, A., Bizoń, B., Cogiel, M., Golba, D., Macioszek, E., & Staniszewski, M. (2023). Development of a realistic crowd simulation environment for fine-grained validation of people tracking methods. *arXiv Preprint* *arXiv:2304.13403*.

Gaggioli, A., Bassi, M., & Fave, A. (2003). Quality of experience in virtual environments. *Emerging Communication*, *5*, 121–136.

Garcia, A., Gomez, F. A., Llorca, C., & Angel-Domenech, A. (2015). Effect of width and boundary conditions on meeting maneuvers on two-way separated cycle tracks. *Accident Analysis & Prevention*, *78*, 127–137.

Geris, A., Cukurbasi, B., Kilinc, M., & Teke, O. (2024). Balancing performance and comfort in virtual reality: A study of FPS, latency, and batch values. *Software: Practice and Experience*, *54*(12), 2336–2348.

Green, M. (2000). " How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors*, *2*(3), 195–216.

Guo, X., Angulo, A., Robartes, E., Chen, T. D., & Heydarian, A. (2022). Orclsim: A system architecture for studying bicyclist and pedestrian physiological behavior through immersive virtual environments. *Journal of Advanced Transportation*, *2022*(1), 2750369.

Hamilton-Baillie, B. (2008). Shared space: Reconciling people, places and traffic. *Built Environment*, *34*(2), 161–181.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*(5), 4282.

Hoogendoorn, S. P., & Bovy, P. H. (2004). Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, *38*(2), 169–190.

Huang, J., Cui, Y., Zhang, L., Tong, W., Shi, Y., & Liu, Z. (2022). An Overview of Agent-Based Models for Transport Simulation and Analysis. *Journal of Advanced Transportation*, *2022*, 1–17.

https://doi.org/10.1155/2022/1252534

Huang, W., Guo, Y., Guo, C., Tang, F., Zhao, Y., Xia, Z., & Zhang, R. (2023). Simulation of urban non-motorized traffic: A agent-based modeling approach based on big data of bike sharing and social force model. *Transactions in Urban Data, Science, and Technology*, *2*(4), 204–222. https://doi.org/10.1177/27541231231180989

Huang, Y., Guo, Z., Chu, H., & Sengupta, R. (2023). Evacuation Simulation Implemented by ABM-BIM of Unity in Students' Dormitory Based on Delay Time. *ISPRS International Journal of Geo-Information*, *12*(4), 160. https://doi.org/10.3390/ijgi12040160

Ihssian, A., & Ismail, K. (2023). Modelling pedestrian safety at urban intersections using user perception. *Accident Analysis & Prevention*, *180*, 106912. https://doi.org/10.1016/j.aap.2022.106912

Janzen, B. F., & Teather, R. J. (2014). Is 60 FPS better than 30?: The impact of frame rate and latency on moving target selection. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 1477–1482. https://doi.org/10.1145/2559206.2581214

Jay, M., Régnier, A., Dasnon, A., Brunet, K., & Pelé, M. (2020). The light is red: Uncertainty behaviours displayed by pedestrians during illegal road crossing. *Accident Analysis & Prevention*, *135*, 105369.

Jiao, Y., Calvert, S. C., Van Cranenburgh, S., & Van Lint, H. (2025). A unified probabilistic approach to traffic conflict detection. *Analytic Methods in Accident Research*, *45*, 100369. https://doi.org/10.1016/j.amar.2024.100369

Johora, F. T., Cheng, H., Müller, J. P., & Sester, M. (2020). *An agent-based model for trajectory modelling in shared spaces: A combination of expert-based and deep learning approaches*.

1878–1880.

Kamal, K., & Farooq, B. (2023). Debiased machine learning for estimating the causal effect of urban traffic on pedestrian crossing behavior. *Transportation Research Record*, *2677*(7), 196–208.

Kaparias, I., Bell, M. G., Biagioli, T., Bellezza, L., & Mount, B. (2015). Behavioural analysis of interactions between pedestrians and vehicles in street designs with elements of shared space. *Transportation Research Part F: Traffic Psychology and Behaviour*, *30*, 115–127.

Kaß, C., Schmidt, G., & Kunde, W. (2018). *Understanding the differences between human and system hazard perception in potential collision situations*. 27–35.

Kenworthy, J., & Newman, P. (2015). *The End of Automobile Dependence: How Cities are Moving Beyond Car-Based Planning*.

Khademi, N., Mazloum, S., Zabihpour, A., & Chen, A. (2024). Designing safer intersections: Exploring the impact of visual and auditory warnings on pedestrian behavior in a virtual simulated environment. *Safety Science*, *178*, 106604.

Kim, Y., Aamir, Z., Singh, M., Boorboor, S., Mueller, K., & Kaufman, A. E. (2025a). Explainable XR: Understanding user behaviors of XR environments using LLM-assisted analytics framework. *IEEE Transactions on Visualization and Computer Graphics*.

Kim, Y., Aamir, Z., Singh, M., Boorboor, S., Mueller, K., & Kaufman, A. E. (2025b). Explainable XR: Understanding user behaviors of XR environments using LLM-assisted analytics framework. *IEEE Transactions on Visualization and Computer Graphics*.

Kim, Y. M., & Rhiu, I. (2024). Development of a virtual reality system usability questionnaire (VRSUQ). *Applied Ergonomics*, *119*, 104319.

Koilias, A., Nelson, M. G., Anagnostopoulos, C., & Mousas, C. (2020). Immersive walking in a virtual crowd: The effects of the density, speed, and direction of a virtual crowd on human movement behavior. *Computer Animation and Virtual Worlds*, *31*(6), e1928.

Kouskoulis, G., Spyropoulou, I., & Antoniou, C. (2018). Pedestrian simulation: Theoretical models vs. Data driven techniques. *International Journal of Transportation Science and Technology*, *7*(4), 241–253.

Kretz, T., Grünebohm, A., Kaufman, M., Mazur, F., & Schreckenberg, M. (2006). Experimental study of pedestrian counterflow in a corridor. *Journal of Statistical Mechanics: Theory and Experiment*, *2006*(10), P10001.

Krumm, J. (2010). *Ubiquitous computing fundamentals*. CRC Press.

Landis, B. W., Petritsch, T. A., Huang, H. F., & Do, A. H. (2004). Characteristics of emerging road and trail users and their safety. *Transportation Research Record*, *1878*(1), 131–139.

Liang, J., Jiang, L., Murphy, K., Yu, T., & Hauptmann, A. (2020). *The garden of forking paths: Towards multi-future trajectory prediction*. 10508–10518.

Linnekamp, K. (2020a). *Modelling Shared Space: An Organized Chaos*.

Linnekamp, K. (2020b). *Modelling Shared Space: An Organized Chaos* [Master's Thesis]. https://studenttheses.uu.nl/handle/20.500.12932/37880

Ljubović, V. (2009). *Traffic simulation using agent-based models*. 1–6.

Macal, C. M., & North, M. J. (2005a). *Tutorial on agent-based modeling and simulation*. 14-pp.

Macal, C. M., & North, M. J. (2005b). *Tutorial on agent-based modeling and simulation*. 14-pp.

Mastora, C., Paschalidis, E., Nikiforiadis, A., & Basbas, S. (2023). Pedestrian Crossings as a Means of Reducing Conflicts between Cyclists and Pedestrians in Shared Spaces. *Sustainability*, *15*(12), 9377. https://doi.org/10.3390/su15129377

Methorst, R., Gerlach, J., Boenke, D., & Leven, J. (2007). Shared space: Safe or dangerous. *A Contribution to Objectification of a Popular Design Philosophy*, *3*.

Michalík, D., Jirgl, M., Arm, J., & Fiedler, P. (2021). Developing an unreal engine 4-based vehicle driving simulator applicable in driver behavior analysis—A technical perspective. *Safety*, *7*(2), 25.

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In *Human behavior and traffic safety* (pp. 485–524). Springer.

Monti, C., Pangallo, M., De Francisci Morales, G., & Bonchi, F. (2023). On learning agent-based models from data. *Scientific Reports*, *13*(1), 9268.

Moody, S., & Melia, S. (2014a). Shared space – research, policy and problems. *Proceedings of the Institution of Civil Engineers - Transport*, *167*(6), 384–392. https://doi.org/10.1680/tran.12.00047

Moody, S., & Melia, S. (2014b). *Shared space–research, policy and problems*. *167*(6), 384–392.

Moreno, A. C., Moreno, M., Porras, C., & Pavón, J. (2023). Human and environmental factors analysis in traffic using agent-based simulation. *Applied Sciences*, *13*(6), 3499.

Motamedi, A., Wang, Z., Yabuki, N., Fukuda, T., & Michikawa, T. (2017). Signage visibility analysis and optimization system using BIM-enabled virtual reality (VR) environments. *Advanced Engineering Informatics*, *32*, 248–262.

Moussaïd, M., Helbing, D., & Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, *108*(17), 6884–6888. https://doi.org/10.1073/pnas.1016507108

Mukoya, K., Weng, E., Choudhury, R., & Kitani, K. (2024). *JaywalkerVR: A VR System for Collecting*

Safety-Critical Pedestrian-Vehicle Interactions. 9600–9607.

Murray, A. (2013). *Information technology law: The law and society*. Oxford University Press, USA.

Nazemi, M., van Eggermond, M. A., Erath, A., & Axhausen, K. W. (2018). Studying cyclists' behavior in a non-naturalistic experiment utilizing cycling simulator with immersive virtual reality. *Arbeitsberichte Verkehrs-Und Raumplanung*, *1383*.

Neo, J. R. J., Won, A. S., & Shepley, M. M. (2021). Designing immersive virtual environments for human behavior research. *Frontiers in Virtual Reality*, *2*, 603750.

Nielsen, J. (2000). *Why you only need to test with 5 users*. Useit. com Alertbox. http://gruponng.com.br/wp-content/uploads/2025/02/Why-You-Only-Need-to-Test-with-5-Users.pdf

Ning, Q., & Li, M. (2022). Modeling Pedestrian Detour Behavior By-Passing Conflict Areas. *Sustainability*, *14*(24), 16522. https://doi.org/10.3390/su142416522

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417.

Payzan-LeNestour, E., Pradier, L., Doran, J., Nave, G., & Balleine, B. (2021). Impact of ambient sound on risk perception in humans: Neuroeconomic investigations. *Scientific Reports*, *11*(1), 5392.

Piaseczna, N., Doniec, R., Sieciński, S., Barańska, K., Jędrychowski, M., & Grzegorzek, M. (2024). Driving Reality vs. Simulator: Data Distinctions. *Electronics*, *13*(14), 2708.

Pöhlmann, K. M., Li, G., Wilson, G., McGill, M., Pollick, F., & Brewster, S. (2024). Is video gaming a cure for cybersickness? Gamers experience less cybersickness than non-gamers in a VR self-motion task. *IEEE Transactions on Visualization and Computer Graphics*.

Quimby, A., & Castle, J. A. (2006). *A review of simplified streetscape schemes*. TRL Limited London.

Raaen, K., & Kjellmo, I. (2015). *Measuring latency in virtual reality systems*. 457–462.

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017). *Are they going to cross? A benchmark dataset and baseline for pedestrian crosswalk behavior*. 206–213.

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health*, *14*(8), 1101–1108.

Sayin, E., Krishna, A., Ardelet, C., Decré, G. B., & Goudey, A. (2015). "Sound and safe": The effect of ambient sound on the perceived safety of public spaces. *International Journal of Research in Marketing*, *32*(4), 343–353.

Seekhao, N., Shung, C., JaJa, J., Mongeau, L., & Li-Jessen, N. Y. (2016). *Real-time agent-based modeling simulation with in-situ visualization of complex biological systems: A case study on vocal fold inflammation and healing*. 463–472.

Seinfeld, S., Schmidt, I., & Müller, J. (2022). Evoking realistic affective touch experiences in virtual reality. *arXiv Preprint arXiv:2202.13389*.

Serino, A., Noel, J.-P., Galli, G., Canzoneri, E., Marmaroli, P., Lissek, H., & Blanke, O. (2015). Body part-centered and full body-centered peripersonal space representations. *Scientific Reports*, *5*(1), 18603.

Serino, A., Noel, J.-P., Mange, R., Canzoneri, E., Pellencin, E., Ruiz, J. B., Bernasconi, F., Blanke, O., & Herbelin, B. (2018). Peripersonal space: An index of multisensory body–environment interactions in real, virtual, and mixed realities. *Frontiers in ICT*, *4*, 31.

Shaaban, K., & Abdelwarith, K. (2020). Pedestrian attribute analysis using agent-based modeling.

Applied Sciences, *10*(14), 4882.

Sieben, A., Schumann, J., & Seyfried, A. (2017). Collective phenomena in crowds—Where pedestrian dynamics need social psychology. *PLoS One*, *12*(6), e0177328.

Silvera, G., Biswas, A., & Admoni, H. (2022). *Dreye vr: Democratizing virtual reality driving simulation for behavioural & interaction research*. 639–643.

Simpson, M. (2020). *Scale and space: Representations in immersive virtual reality*. The Pennsylvania State University.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557.

Soares, F., Silva, E., Pereira, F., Silva, C., Sousa, E., & Freitas, E. (2021). To cross or not to cross: Impact of visual and auditory cues on pedestrians' crossing decision-making. *Transportation Research Part F: Traffic Psychology and Behaviour*, *82*, 202–220.

Sulle, M., Mwakalonge, J., Comert, G., Siuhi, S., Gyimah, N. K., Roberts, J., & Ruganuza, D. (2025). Analysis of Distracted Pedestrians Crossing Behavior: An Immersive Virtual Reality Application. *arXiv Preprint arXiv:2503.16443*.

Taillandier, P., Gaudou, B., Grignard, A., Huynh, Q.-N., Marilleau, N., Caillou, P., Philippon, D., & Drogoul, A. (2019). Building, composing and experimenting complex spatial models with the GAMA platform. *GeoInformatica*, *23*, 299–322.

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *International Journal of Virtual Reality*, *16*(1), 33–48.

Tzouras, P. G., Batista, M., Kepaptsoglou, K., Vlahogianni, E. I., & Friedrich, B. (2023). Can we all coexist? An empirical analysis of drivers' and pedestrians' behavior in four different shared space road environments. *Cities*, *141*, 104477.

Ugwitz, P., Šašinková, A., Šašinka, Č., Stachoň, Z., & Juřík, V. (2021). Toggle toolkit: A tool for conducting experiments in unity virtual environments. *Behavior Research Methods*, *53*(4), 1581–1591.

Unity Asset Store. (n.d.). Retrieved June 3, 2025, from https://assetstore.unity.com

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology*, *18*, 1–18.

Wang, J., Shi, R., Zheng, W., Xie, W., Kao, D., & Liang, H.-N. (2023). Effect of frame rate on user experience, performance, and simulator sickness in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, *29*(5), 2478–2488.

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. MIT press.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225–240.

Yuan, Y., Daamen, W., Goñi-Ros, B., & Hoogendoorn, S. P. (2018). Investigating cyclist interaction behavior through a controlled laboratory experiment. *Journal of Transport and Land Use*, *11*(1), 833–847.

Zhang, C. (2020). *Investigation on motion sickness in virtual reality environment from the perspective of user experience*. 393–396.

Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PloS One*, *11*(6), e0157795.

Zhao, J., LaFemina, P., Carr, J., Sajjadi, P., Wallgrün, J. O., & Klippel, A. (2020). *Learning in the field: Comparison of desktop, immersive virtual reality, and actual field trips for place-based STEM education*. 893–902.

Zhao, J., Wallgrün, J. O., Sajjadi, P., LaFemina, P., Lim, K. Y., Springer, J. P., & Klippel, A. (2022). Longitudinal effects in the effectiveness of educational virtual field trips. *Journal of Educational Computing Research*, *60*(4), 1008–1034.

Zhu, M., Li, H., Sze, N. N., & Ren, G. (2022). Exploring the impacts of street layout on the frequency of pedestrian crashes: A micro-level study. *Journal of Safety Research*, *81*, 91–100. https://doi.org/10.1016/j.jsr.2022.01.009

Appendix

Appendix 1. Links to Gihub Repository

git@github.com:triplexgu/MGI_ABM_vr.git

Appendix 2. Metrics logged during simulation

Logged csv	Metrics	Description	
MovementData.csv	*Time*	Current timestamp (seconds)	Time.time
	PositionX/Y/Z	Player's current position	cameraRig.centerEyeAnchor.position
	Player_spd	Player's current speed (km/h)	The speed calculated using the previous frame position and the current frame position is converted to km/h
	ph_conflictProb	between the player and the nearest agent The probability of physical collision	Considering physical size, heading difference and relative speed, *ComputeConflictProbability()* calculate
	ph_edgeDistance	Physical edge distance between the player and the agent (m)	Center distance - Radius of both sides (using GetVisualRadius()）
	vs_conflictProb	between the player and the nearest agent The probability of visual conflict	Add psychological redundancy (+0.3m) to the physical radius and then call *ComputeConflictProbability()*
	vs_edgeDistance	The visual edge distance between the player and the agent (m)	Similar to above, but with 0.3m added
	GapSize_sec	The player pause time (seconds) , indicating the gap duration	*GapSize_sec* = Time.time - pauseStartTime
CollisionData.csv	*Time*	Current timestamp (seconds)	Same as MovementData.csv
	PositionX/Y/Z	Player's current position	Same as MovementData.csv
	AgentTag	The closest agent type (Pedestrian, Cyclist, Moped)	pass nearestAgent.tag
	ph_conflictProb	Probability of physical conflict when conflict is triggered	Same as MovementData.csv

	vs_conflictProb	The probability of visual conflict when a conflict is triggered	Same as MovementData.csv
	UserResponse	Whether the user confirms that this is a conflict (Yes/No/0)	Yes / No – options of Unity UI panel 0 – participants press “A” to record collision manually *HandleUserResponse()*
PerformanceData.csv	*Time*	Current timestamp (seconds)	Same as MovementData.csv
	FPS	Frames per second (recorded every 1 second)	Unity 1.0f / Time.unscaledDeltaTime, which indicates the countdown of the time taken for this frame. *FPS* = Time.time - lastFpsUpdateTime >= 1.0f Record once
	Latency(ms)	Latency (time per frame), calculated from FPS: 1000 / avgFPS	Calculated from FPS: Latency = 1000 / FPS

Appendix 3. Recording of Training session

https://youtu.be/f1HBGCK1dnM

Appendix 4. Recording of Simulation session

https://youtu.be/AD7txzpwgXo

Appendix 5. Link to post-simulation questionnaire

https://wur.pdx1.qualtrics.com/homepage/ui

Appendix 6. Descriptive statistics of questionnaire’s quantitative statements

Statements	**Mean**	**Median**	**Std**
Presence			
1. My interactions with the virtual environment seemed natural.	6.2	6.5	1.87
2. The devices (hand-held controllers) which controlled my movement in the virtual environment seemed natural.	6.4	7	1.71
3. I was able to actively survey the virtual environment using vision.	7.9	8	1.29
4. I had the feeling that I was in the middle of the action rather than merely observing.	8.1	8	1.29
5. I felt like I was actually there in the environment	7.3	7	1.83

of the presentation.			
6. It was as though my true location had shifted into the environment in the presentation.	7.4	7.5	1.84
7. I felt as though I was physically present in the environment of the presentation.	7	7	1.83
8. I had the impression that I could act in the environment of the presentation.	7.1	7.5	1.66
9. I felt like I could move around among the objects in the presentation.	7.8	8	1.32
10. It seemed to me that I could have some effect on things in the presentation, as I do in real life.	5.7	6.5	2.26
Usability			
1. I think that I would like to use this system frequently.	6.3	7	2.54
2. I found the system unnecessarily complex.	7.9	8	1.1
3. I would imagine that most people would learn to use this system very quickly.	7.9	8	1.29
4. I needed to learn a lot of things before I could get going with this system.	8.8	9	1.32
5. If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me.	8.1	8.5	1.73
6. Using the interaction devices (VR headset and controllers) is a bad idea.	9.6	10	0.52
Emotion			
1. I enjoy using this technology.	8.2	8.5	2.7
2. At each step, I knew what to do.	7.3	8	2.45
3. I felt I controlled the situation.	7.5	8	2.51
4. I felt I was experiencing an exciting moment.	7.6	8	2.5
5. I enjoyed being in this virtual environment.	7.9	8	2.64
6. I felt nervous in the virtual environment.	6.4	7	2.8
7. I found my mind wandering while I was in the virtual environment.	6.7	7	2.95
8. The interaction devices (VR headset and controllers) bored me to death.	9.2	9	0.63
9. I enjoyed dealing with the interaction devices (VR headset and controllers).	7	8	3.06
Cybersickness			
1. I suffered from fatigue during my interaction with the virtual environment.	4.7	6	2.83
2. I suffered from headache during my interaction with the virtual environment.	3.7	2	3.09
3. I suffered from eyestrain during my interaction	4.8	4.5	2.97

with the virtual environment.			
4. I felt an increase in my salivation during my interaction with the virtual environment.	3.3	3	1.42
5. I suffered from nausea during my interaction with the virtual environment.	3.8	2	3.49
6. I suffered from fullness of the head during my interaction with the virtual environment.	5.7	7.5	3.71
7. I suffered from dizziness with eyes open during my interaction with the virtual environment.	5.2	5	3.52
8. I suffered from vertigo during my interaction with the virtual environment.	5	4.5	3.68

Appendix 7. Summary of users' open-end feedbacks

Questions	**Themes**	**Code examples**	**Frequency**
1. What were your favorite parts of this application?	Interaction	• Interaction with the real environment through VR. • The interactive aspect was quite nice.	4 (n = 40%)
	Useful tutorial	• The tutorial part is clear.	1 (n = 10%)
	Realistic environment modelling	• Real environment and scene. • The 3D model looks nice.	3 (n = 30%)
2. What were the biggest challenges or frustrations you faced while using this application?	Control difficulty	• The interaction is not very regular as traditional controllers need to get used to it • It is difficult to coordinate with the controllers while moving. • I feel that the switch is less very user-friendly, because I have to find out the button based on my feelings.	3 (n = 30%)
	Cybersickness	• dizziness with eyes • I felt dizzy and uncomfortable during this VR simulation • Motion sickness!!!	4 (n = 40%)
3. Do you have any suggestions for additional functions or elements that could make this application more useful or enjoyable for you?	Environmental Variation	• The dummies should not spawn altogether at the same places. • Adding more character models or bike colors can drastically change the behavior of the participant.	2 (n = 20%)
	Multi-sensory Enhancement	• Sounds can be added to simulate the actual environment.	6 (n = 60%)

		• Interaction with collision events such as shake feedback. • Could add some background music to make this VR environment more immersive. • Maybe a vibration could be added when there is a collision. • Adding spatial sound (bike horns, water sound etc.) will make the environment more enjoyable to be immersed in.	
4. Imagine you are a researcher using this application to collect human behavioral data for improving transport simulations. Do you find the visualizations easy to interpret? If not, what additional metrics or types of information would make them more useful to you?	Need for Clearer Feedback and Metrics	• I didn't get collided as feedback with some words hints. • It would be nice to have a clear final value written for those who are not in the field.	2 (n = 20%)
	Easy Interpretation	• I think it can be visualized easily. • It's explainable for this visualization. • They seemed pretty easy to interpret. • Simulations like these are quite popular for people to follow through and easy to record different metrics regarding the transportational behavior of the player.	6 (n = 60%)