Social-Aware Autonomous Vehicle-Pedestrian Coupled Decision-Making and Behavior Planning

Meiting Dang

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

JAMES WATT SCHOOL OF ENGINEERING

COLLEGE OF SCIENCE AND ENGINEERING

UNIVERSITY OF GLASGOW

09 2025

Declaration

I certify that the thesis presented here for examination for a PhD degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Meiting Dang

Abstract

Rapid advancements in vehicle automation technology have enabled autonomous vehicles (AVs) to move beyond simple, structured highway settings and operate increasingly within more complex urban environments. As AVs become more prevalent, they will inevitably share the road with diverse traffic participants, particularly pedestrians. The high uncertainty and dynamic nature inherent in human behavior significantly complicate the AV's decision-making process. To ensure both safety and efficiency while exhibiting socially acceptable behavior, AVs must make real-time decisions and continuously adapt their strategies in response to surrounding pedestrian behaviors. This poses a major challenge to existing AV decision-making systems. Therefore, this study focuses on developing the decision-making strategies for AVs in urban pedestrian-involved environments, spanning scenarios from simple interactions with a single pedestrian to complex cases involving multiple pedestrians. The core objective is to uncover the underlying interaction dynamics and to improve the AV's capability to make decisions that are safe, efficient and socially acceptable in dynamic, human-centric contexts.

Initially, this study investigates the critical factors influencing the decision-making processes of human drivers and pedestrians during vehicle–pedestrian interactions. A series of controlled experiments were conducted using a virtual reality platform designed to collect realistic behavioral data. The data analysis mainly focused on kinematic variables that are easily measurable in real-world deployments. AdaBoost classification and Partial Dependence Plots were employed to identify and visualize the most influential factors affecting pedestrian crossing intentions and driver approaching behaviors. The results indicate that longitudinal distance and vehicle acceleration are the most influential factors in pedestrian decision-making, while pedestrian speed and longitudinal distance also play a crucial role in determining whether the vehicle yields or not. Based on these insights, a simplified mathematical model was developed to relate observable kinematic parameters to pedestrian crossing intentions, providing a practical tool for dynamically inferring crossing intentions during interactions. Additionally, the study explored driver yielding patterns under varying degrees of pedestrian intention clarity. These findings support the development of interpretable and implementable models for real-time decision-making in vehicle–pedestrian interaction scenarios.

In addition, this study examines the decision-making strategies of the AVs when interacting with a single pedestrian in unsignalized intersections. A novel framework was proposed that integrates the Partially Observable Markov Decision Process with behavioral game theory to dynamically model interactive behaviors between the AVs and the

pedestrian. Both agents were modeled as dynamic-belief-induced quantal cognitive hierarchy models, considering human reasoning limitations and bounded rationality in the decision-making process. Moreover, a dynamic belief updating mechanism allowed the AV to update its understanding of the opponent's rationality degree in real-time based on observed behaviors and adapt its strategies accordingly. The analysis results indicate that proposed models effectively simulate vehicle-pedestrian interactions and the AV decision-making approach performs well in safety, efficiency, and smoothness. It captures key patterns of the driving behavior operated by real human drivers in virtual reality experiments and even achieves more comfortable navigation compared to our previous virtual reality experimental data.

Finally, this study investigates the decision-making strategies of the AVs operating in pedestrian-rich shared spaces. A novel framework was proposed for modeling interactions between the AV and multiple pedestrians. In this framework, a cognitive process modeling approach inspired by the Free Energy Principle was integrated into both the AV and pedestrian models to simulate more realistic interaction dynamics. Specifically, the proposed pedestrian Cognitive-Risk Social Force Model adjusts goal-directed and repulsive forces using a fused measure of cognitive uncertainty and physical risk to produce human-like trajectories. Meanwhile, the AV leverages this fused risk to construct a dynamic, risk-aware adjacency matrix for a Graph Convolutional Network within a Soft Actor-Critic architecture, allowing it to make more reasonable and informed decisions. The qualitative and quantitative results indicate that the proposed strategy effectively improves safety, efficiency, and smoothness of AV navigation compared to the state-of-the-art method.

In conclusion, this study addresses key challenges in AV decision-making within human-centric urban environments, providing valuable insights as well as practical solutions to support safe, efficient, and socially aware autonomous navigation.

List of Publications

- **M Dang**, Y Jin, P Hang, L Crosato, Y Sun, C Wei. Coupling intention and actions of vehicle–pedestrian interaction: A virtual reality experiment study[J]. *Accident Analysis & Prevention*, 2024, 203: 107639.

- **M Dang**, D Zhao, Y Wang, C Wei. Dynamic Game-Theoretical Decision-Making Framework for Vehicle-Pedestrian Interaction With Human Bounded Rationality[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025.

- **M Dang**, Y Wu, Y Wang, D Zhao, D Flynn, C Wei. Free Energy-Inspired Cognitive Risk Integration for AV Navigation in Pedestrian-Rich Environments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, under review, 2025.

- Y Wu, **M Dang**, L Wu, E Ho, Z Chen, C Wei. ASINet: Energy-Based Adaptive Framework for Spatiotemporal Interaction-Aware Pedestrian Intention Prediction. *IEEE Transactions on Neural Networks and Learning Systems*, under review, 2025.

- C Wei, K Tian, W Lyu, **M Dang**. Enabling Safe AV–Pedestrian Interactions Through Implicit Decision-Making and Explicit eHMI Communication. In: *Comprehensive Robotics for Extreme and Challenging Environments*. Elsevier, under review, 2025.

Contents

List of Tables

List of Figures

Abbreviations

A3C	Advantage Actor-Critic
AdaBoost	Adaptive Boosting
ADE	Average Displacement Error
AV	Autonomous Vehicle
BC	Behavioral Cloning
CA	Cellular Automaton
CH	Cognitive Hierarchy
CIL	Conditonal Imitation Learning
CNN	Convolutional Neural Network
CR	Collision Rate
CR-SFM	Cognitive-Risk Social Force Model
CTCE	Centralized Training Centralized Execution
CTDE	Centralized Training Decentralized Execution
CV	Constant Velocity
DARPA	Defense Advanced Research Projects Agency
DB-QCH	Dynamic-Belief-Induced Quantal Cognitive Hierarchy Model
DDPG	Deep Deterministic Policy Gradient
DDQN	Double Deep Q-Learning
DQN	Deep Q-Learning
DRL	Deep Reinforcement Learning
DTDE	Decentralized Training Decentralized Execution
FDE	Final Displacement Error
FEP	Free Energy Principle
FSM	Finite State Machine
GAIL	Generative Adversarial Imitation Learning
GAN	Generative Adversarial Network
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
G-SAC-Cog	Graph-Enhanced SAC with Cognitive Modeling
G-SAC-NoCog	Graph-Enhanced SAC without Cognitive Modeling

HBS	Hamburg Bergedorf Station
IL	Imitation Learning
IRL	Inverse Reinforcement Learning
KL	Kullback-Leibler
LSTM	Long Short-Term Memory
MADDPG	Multi-Agent Deep Deterministic Policy Gradient
MAPPO	Multi-Agent Proximal Policy Optimization
MARL	Multi-Agent Reinforcement Learning
MASAC	Multi-Agent Soft Actor-Critic
MCTS	Monte Carlo Tree Search
MDP	Markov Decision Process
MLP	Multilayer Perceptron
MPC	Model Predictive Control
ORCA	Optimal Reciprocal Collision Avoidance
PDP	Partial Dependence Plot
POMDP	Partially Observable Markov Decision Process
PPO	Proximal Policy Optimization
QR	Quantal Response
RA-SFM	Risk-Aware Social Force Model
RL	Reinforcement Learning
RNN	Recurrent Neural Network
RQ	Research Question
SAC	Soft Actor-Critic
SAE	Society of Automotive Engineering
SFM	Social Force Model
S-SAC	Standard Soft Actor-Critic
SVO	Social Value Orientation
TD3	Twin Delayed Deep Determinsitic Policy Gradient
TTA	Time-to-Arrival
TTC	Time-to-Collision
UAW-PCG	Uncertainty-Aware Polar Collision Grid
UCT	Upper Confidence Bound for Trees
UE	Unreal Engine
VO	Velocity Obstacle
VR	Virtual Reality

Acknowledgements

Completing a PhD is a long and challenging journey. Throughout the years of study, I have not only gained academic knowledge and research experience, but also grown personally through moments of confusion, perseverance, and reflection. Now that this journey is approaching its end, I would like to express my heartfelt gratitude to all those who have supported, encouraged, and accompanied me along the way.

First and foremost, I am deeply grateful to my supervisor, Dr. Chongfeng Wei, for giving me the opportunity to pursue my PhD in the UK and for his invaluable guidance throughout the course of my research. His academic insight, rigorous standards, and patient supervision helped me find direction when I was uncertain, and gave me confidence when I made progress. His support has been instrumental in shaping both my research and my academic mindset.

I would also like to express my sincere thanks to my husband, Mr. Feng Gao, for his unwavering love and understanding. He has stood by me through all the pressures and emotional ups and downs of this journey, always offering steadfast encouragement and quiet strength. His presence has been a constant source of comfort and support throughout my PhD.

I am especially thankful to my parents and my brother for their unconditional love and endless support. They have accompanied me through every step of my education with patience and care. I owe everything I have achieved today to their sacrifices, encouragement, and unwavering belief in me. Their strength has been the foundation of my own.

I would also like to sincerely thank my second supervisors, Dr. Dezong Zhao and Prof. David Flynn, for their support and constructive feedback on my thesis. Your detailed comments, patient revisions, and rigorous academic guidance have helped me significantly improve the quality of my work. I truly appreciate your time, effort, and mentorship.

Finally, I am grateful to everyone who has supported me academically, emotionally, or otherwise throughout this doctoral journey. Every kind word, every gesture of encouragement, and every moment of understanding has been a light along my path.

To all of you—thank you from the bottom of my heart.

Chapter 1

Introduction

Autonomous vehicles (AVs) are equipped with advanced sensors, artificial intelligence, and computing systems, enabling them to navigate and operate independently without direct human control [1]. The advent of this technology signifies a substantial stride in the automotive sector, promising a future of transportation in a safer, more efficient, and more convenient manner. Specifically, self-driving cars are expected to significantly reduce traffic accidents because they can respond more quickly than human drivers, thereby reducing the risk of human error. In addition, they can optimize travel routes, avoid traffic congestion, and shorten travel times. This innovation also enhances accessibility for individuals unable to drive due to age, disability, or other limitations.

The Society of Automotive Engineering (SAE) has identified six levels of vehicle automation, varying from no automation to full automation [2]. At level 0, there is no automatic support and the human driver retains full control over the vehicle's operation and navigation without any automated assistance. In contrast, the vehicle reduces its reliance on the human driver and independently performs more driving tasks at higher automation levels. These levels involve specific tasks such as cruise control and steering, corresponding to levels 1 and 2, respectively. At levels 3 and 4, the vehicle can manage the majority of driving tasks under certain conditions but requires human intervention if requested by the system. The highest-level vehicle is anticipated to be entirely autonomous and can complete all driving activities without the assistance of a human operator.

The pursuit of achieving full automation in the field of AVs has led numerous researchers to focus on multiple aspects of AV technology. This technology encompasses a complex system involving perception, decision-making, planning, and control [3], allowing vehicles to navigate and operate without human participation. The perception module collects data about the vehicle's surroundings through an array of sensors, enabling the vehicle to perceive its environment. Once the vehicle has a thorough understanding of its surroundings, the algorithm will assess the situation and determine the best course of action. Upon

decision-making, the vehicle will devise a safe and efficient navigation route to the destination. Subsequently, the control system, following the produced decision and planned path, will precisely control the vehicle's movements by sending signals to the vehicle's actuators, which include steering, braking, and acceleration systems. As improvements in AV technology have facilitated the integration of AVs into practical applications, there has been a growing deployment of AV in the real world.

In recent years, rapid advancements in vehicle automation technology have enabled AVs to expand beyond simple, structured highway environments and to be increasingly deployed in more complex urban environments [4]. They are expected to become a key component of future urban transportation systems [5]. As AVs become widespread, they will inevitably share the road with various road users, particularly pedestrians [6], [7].

Unlike vehicles that follow well-defined rules and predictable physical dynamics, pedestrians exhibit a high degree of behavioral variability, uncertainty, and unpredictability [8]. They may change direction abruptly, hesitate at crossings, or even ignore traffic signals. In addition, pedestrian behavior is often shaped by subjective and context-dependent factors, such as intent, attention, urgency, or social influence, which are difficult for AVs to directly perceive or infer.

This challenge is further amplified in urban environments, where interactions between AVs and pedestrians are frequent, dynamic, and socially embedded. AVs must make real-time decisions and continuously adapt their strategies in response to surrounding pedestrians. Unlike highway driving, urban navigation often involves implicit negotiations, such as eye contact, body language, or yielding gestures, which are hard to quantify algorithmically but play a crucial role in ensuring smooth and safe interaction.

(a)

(b)

Figure 1.1: Two representative AV pedestrian interaction scenarios investigated in this thesis. (a) Interaction between an AV and a single pedestrian at an unsignalized intersection (from [9]). (b) Interaction between an AV and multiple pedestrians in an urban shared space (from [10]).

These human-specific traits introduce significant complexity to AV decision-making and motion planning, as AVs must not only detect pedestrians accurately, but also interpret their intentions and adapt their own actions accordingly. Ensuring safe, efficient, and socially compliant interactions with pedestrians has become one of the core challenges in the deployment of AVs in human-centric environments. Therefore, this study focuses on AV-pedestrian interactions and aims to develop AV decision-making frameworks that tackle these issues in pedestrian-involved environments. Fig. 1.1 illustrates the two types of scenarios studied in this work.

1.1 Challenges for AV-Pedestrian Interaction

Despite advances in autonomous navigation, AVs still face several challenges when interacting with pedestrians in complex urban environments. This section identifies three major challenges in AV-pedestrian interaction.

Challenge 1: Modeling Dynamic Mutual Interaction Between the AV and Pedestrians. One of the fundamental challenges in AV-pedestrian interaction lies in constructing models that can accurately capture the mutual and dynamic interplay between autonomous vehicles and pedestrians. This challenge arises from the fact that their interactions are inherently interdependent and tightly coupled. Each agent is required not only to make its own decisions, but also to adapt its behavior in real time based on the evolving actions of the other party. For example, pedestrians continuously adjust their behavior, such as walking speed, trajectory or crossing decisions in response to the perceived motion and intentions of the AV, while the AV must respond promptly to pedestrians' often unpredictable and abrupt behavioral changes. This results in a complex feedback loop of continuous reciprocal influence, which is far more intricate than static obstacle avoidance or unidirectional interaction scenarios. Developing accurate and adaptive models of this real-time interaction is crucial for gaining a deeper understanding of AV-pedestrian interaction mechanisms. Such models can also significantly improve the development of autonomous navigation strategies, achieving safe and naturalistic AV navigation in complex urban environments.

Challenge 2: Balancing Safety and Efficiency in AV Decision-Making. Another critical challenge in AV-pedestrian interaction is achieving an effective balance between safety and driving efficiency in AV's decision-making process. AVs are obligated to ensure pedestrian safety at all times and thus often adopt overly conservative strategies, such as frequent stops or unnecessary deceleration, or yielding even in ambiguous situations, when approaching or interacting with pedestrians. While these behaviors

reduce the risk of collision, they may also lead to increased travel time and disrupted traffic flow. Moreover, excessive caution can result in unnatural or passive behavioral responses that differ from those of human drivers. Such behaviors may confuse pedestrians or lead to hesitation during interactions, ultimately impacting public trust and social acceptance of AVs. Conversely, prioritizing efficiency without sufficient consideration of safety may cause high-risk situations with potentially life-threatening consequences. Therefore, striking the right balance between safety and efficiency remains both technically challenging and essential for real-world deployment, especially in dynamic pedestrian-involved environments.

Challenge 3: The Interpretability of AV Decision-Making. A third important challenge is improving the interpretability and transparency of the AV's decision-making process. Many state-of-the-art AV decision-making policies rely on deep reinforcement learning or end-to-end neural architectures, which are often treated as black-box models. This opaque decision-making mechanism makes it difficult to explain why a particular decision was made thereby hindering the analysis, debugging, and improvement of the policy. Moreover, when AVs make decisions that are unnatural, inconsistent, or misaligned with social expectations, pedestrians may find them unintuitive or difficult to interpret. Such unintelligent behaviors often stem from an underlying decision-making policy that lacks transparency, making it difficult for humans to anticipate the vehicle's actions or engage in smooth interactions. This can lead to confusion, misunderstanding, or even unsafe interactions, especially in densely populated urban settings. Addressing this challenge requires the development of decision-making frameworks that incorporate transparent reasoning processes and socially aligned behavioral patterns, enabling AVs to operate in ways that are both safe and intuitively understandable to humans.

In response to these challenges, this study proposes decision-making frameworks for AV-pedestrian interactions, with the goal of enhancing safety, efficiency, and interpretability in socially interactive scenarios.

1.2 Research Gap

This study focuses exclusively on the decision-making process of AVs in their interaction with pedestrians, with interactions involving human-driven vehicles beyond its scope. Although a substantial body of research has examined AV–pedestrian interactions, several critical knowledge gaps remain. This thesis aims to address these gaps by investigating vehicle-pedestrian interactive behaviors and incorporating the resulting insights into interaction modelling and decision-making algorithm development. The key research gaps identified are as follows:

Gap 1: Lack of human-behavior factor analysis suitable for safe AV navigation

Although many existing studies analyze the factors influencing human driver behavior, their primary aim is behavioral explanation rather than supporting safe AV navigation. This is because that these studies often rely on internal or unobservable factors, such as personality traits or emotions, which cannot be measured by an AV in real-world driving when interacting with pedestrians. Consequently, the findings from such works cannot be directly used to develop AV decision-making models, which must rely solely on externally observable and measurable variables. The specific observable factors that influence human driver decision-making, and that are accessible to an AV in practice, have not been clearly identified in prior studies.

Gap 2: Limited exploration of AV decision-making in diverse AV-pedestrian interaction scenarios.

Most existing studies on AV-pedestrian interactions have primarily concentrated on post-hoc behavioral analyses, such as examining video recordings of real-world interactions [9] or conducting simulation-based experiments [11] to describe how pedestrians cross and how vehicles yield. These efforts have provided valuable insights into human behavior and interaction dynamics in traffic environments. However, they largely remain at the level of behavioral explanation and seldom translate these findings into algorithmic frameworks for AV decision-making. As a result, how AVs should act and make decisions across diverse interaction scenarios, ranging from relatively simple single-pedestrian cases to more complex multi-pedestrian shared-space environments, remains insufficiently explored, leaving a critical gap between behavioral understanding and algorithmic implementation.

Gap 3: Oversimplified pedestrian models in AV-pedestrian interaction studies.

In the AV-pedestrian interaction modeling, existing approaches tend to emphasize the AV decision-making process while overlooking the role of pedestrian behavior modeling. As a result, pedestrians are often described using overly simplified representations that treat them as passive entities (e.g., constant velocity [12], fixed trajectories [13]). Such simplifications fail to capture the dynamic, adaptive, and bidirectional nature of real-world interactions, in which pedestrians continuously adjust their behavior in response to AV actions, and AVs in turn adapt to pedestrians. This lack of realism in pedestrian modeling reduces the accuracy of interaction and ultimately limits the capability and reliability of AV decision-making algorithms in complex real-world environments.

Gap 4: Lack of AV decision-making models grounded in human cognition.

AVs are expected to make decisions in a human-like manner, such as reasoning about their environment, anticipating others' actions, and producing responses that align with human behavioral patterns. Yet, most existing AV decision-making methods are developed within traditional optimization [14] or data-driven paradigms [15], which, although effective in generating actions, fail to capture the cognitive reasoning processes underlying human decision-making in interactive scenarios. Incorporating human cognition into AV decision-making therefore represents a new modeling perspective, which would allow AV models to more closely mirror human logic and enhance the interpretability of their decisions. Nevertheless, only limited studies have attempted this integration, and the role of human cognition in AV decision-making remains insufficiently explored, motivating this study to investigate new approaches for integrating cognitive processes into AV–pedestrian interaction modeling.

1.3 Research Problems and Objectives

To address the identified research gaps, the primary objective in this work is to improve the understanding of AV-pedestrian interaction mechanism, to develop interaction models across diverse scenarios, and to enhance the ability of AVs to make decisions that are safe, efficient and socially acceptable in dynamic, human-centric environments. Specifically, the study addresses the following three core research questions (RQs).

RQ1: What are the most critical factors that influence decision-making processes of human drivers and pedestrians, respectively, during their interactions in urban environments and how do these factors shape their behaviors?

This research question seeks to identify the essential factors that impact human decision-making processes in urban scenarios. Understanding these factors is a critical foundation for constructing accurate and human-like models of both the AV and pedestrian. A more realistic pedestrian decision model enables the simulation of more authentic interactions, allowing the AV's decision-making algorithms to be trained and evaluated in environments that reflect real human behavior. At the same time, identifying the key elements of human driver behavior can guide the development of AV policies that mimic human reasoning and reactions. Together, these insights contribute to more socially acceptable AV behavior and facilitate a deeper understanding of the dynamic interaction mechanisms between AVs and pedestrians.

RQ2: In scenarios involving a single pedestrian at unsignalized intersections, how can an AV accurately interpret pedestrian intentions and make appropriate decisions in the absence of explicit traffic signals?

In unsignalized interaction scenarios, the priority of right-of-way is often ambiguous due to the absence of traffic signals, leading to potential road conflicts that further complicate AV's decision making. Pedestrians typically rely on observable cues to make crossing decisions, while the AV must accurately perceive the pedestrian's intention, such as whether they intend to cross or yield, and respond accordingly to ensure safe and efficient interactions. Addressing this problem is not only crucial for reliable AV operation in common urban settings but also for establishing a foundation for handling more complex multi-pedestrian interactions. Investigating the interactions between an AV and a single pedestrian provides a simplified yet meaningful basis for understanding of the core mechanisms underlying human-vehicle negotiation.

RQ3: In complex and dynamic urban shared spaces, how can an AV effectively manage interactions with multiple pedestrians while maintaining safety, efficiency, and social acceptance?

In urban shared space scenarios, AVs are required to simultaneously interact with multiple pedestrians who may exhibit diverse behaviors, intentions, and trajectories. These interactions are inherently uncertain and highly dynamic, as each pedestrian may respond differently to the AV's actions or to other pedestrians nearby. This complexity poses significant challenges for AV decision-making, which must account for the interdependence among agents and react to their possible behaviors in real time. Developing decision-making strategies that can effectively manage these multi-agent interactions is essential for the safe operation of AVs in densely populated, human-centric environments. Moreover, the ability to handle complex interactions with multiple pedestrians is a critical step toward the practical deployment of AVs in real-world urban settings.

Each question corresponds to a key aspect of AV decision-making: understanding human behavior during the decision-making process (RQ1), designing AV responses in simple scenarios (RQ2), and generalizing to complex multi-agent settings (RQ3). These problems are investigated with the aim of improving AV performance in human-centric environments.

Overall, the challenges identified in the previous section point to limitations in existing research and thereby define the corresponding research gaps, which in turn motivate the research questions. Challenge 1 demonstrates the need to understand interactive behavior using information accessible to AVs, leading to Gap 1 and Research Question 1. Challenges 1 and 2 together indicate shortcomings in existing interaction-modeling approaches, giving rise to Gaps 2-4 and motivating Research Questions 2 and 3. Finally, Challenge 3 highlights the need for interpretable AV decision-making, underpinning Gap 4 and providing further motivation for Research Questions 2 and 3.

1.4 Contributions

The research gaps identified in Section 1.2, Chapter 1, motivated the formulation of the research questions outlined in Section 1.3, Chapter 1. To fill these gaps and address the corresponding research questions, this thesis makes the following key contributions:

- Virtual reality experiments were conducted to collect detailed interaction data between the vehicle and a single pedestrian using a state-of-the-art distributed driving–pedestrian simulator platform. This dataset was then used to investigate the critical factors influencing the decision-making processes of human drivers and pedestrians, respectively. In addition, the analysis explored the underlying relationships between motion parameters and pedestrian behaviors, as well as the characteristics of driver-yielding behavior. This contribution directly addresses the first research question and provides a foundational understanding of real-world interactions at unsignalized intersections, which is essential for developing AV decision-making models in complex urban contexts.
- A novel decision-making framework was proposed that integrates the partially observable markov decision process (POMDP) with behavioral game theory to dynamically model interactions between an AV and a single pedestrian at unsignalized intersections. Both the AV and the pedestrian are modeled as dynamic-belief-induced quantal cognitive hierarchy (DB-QCH) agents, considering human reasoning limitations and bounded rationality in the decision-making process. Moreover, a dynamic belief updating mechanism was introduced, allowing the AV to update its under-

standing of the pedestrian's rationality level in real-time based on observed behaviors and to adapt its strategies accordingly. This contribution addresses the second question, and paves the way for the development of explainable and trustworthy AV decision-making systems in interactive, human-involved environments.

- An innovative framework was proposed to tackle the decision-making challenges faced by the AV when navigating complex urban environments involving multiple pedestrians. A cognitive uncertainty modeling approach based on the free energy principle was introduced to emulate human reasoning and quantify cognitive uncertainty during social interactions. Building upon this, a novel pedestrian model, referred to as the Cognitive-Risk Social Force Model (CR-SFM), was developed, which enables pedestrian agents to autonomously adapt their motions in uncertain and interactive conditions. This yields more realistic behaviors compared to traditional models with fixed coefficients, and provides a reliable simulation foundation for training and evaluating AV decision-making strategies. Furthermore, the AV's decision-making policy was trained using the Soft Actor-Critic (SAC) algorithm, where a risk-encoded interaction graph was integrated by redefining the adjacency matrix to encode dynamic interaction risks. This allows the policy to better perceive interactive relationships and potential threats, thereby improving decision-making performance. This contribution directly supports the third research question and provides a potential solution for AV navigation in handling complex, dynamic, and uncertain human-involved scenarios.

1.5 Thesis outline

This thesis is organized into six chapters, each corresponding to a major component of the research and collectively addressing the three core research questions.

Chapter 1 introduces the background and motivation for the research, highlights the key challenges in AV–pedestrian interactions, and defines the overall research direction and scopes. It further identifies the existing research gaps. Based on these gaps, this chapter formulates the central research questions, summarizes the main contributions, and presents the structure of the thesis.

Chapter 2 provides a comprehensive review of related work in the field, covering four main sections: (1) interaction data collection methods, (2) human behavior analysis, (3) pedestrian behavior modeling, and (4) AV decision-making approaches.

Chapter 3 addresses the first research question by presenting the design and implementation of VR-based experiments to simulate AV–pedestrian interactions. It also details the data collection and analysis methods.

Chapter 4 introduces a decision-making framework for the AV interacting with a single pedestrian at unsignalized intersections, responding to the second research question.

Chapter 5 tackles the third research question, focusing on AV decision-making in complex urban environments involving multiple pedestrians.

The final chapter concludes the thesis by summarizing the key findings and contributions. It also discusses the limitations of the current work and outlines potential directions for future research.

Chapter 2

Literature Review

This chapter provides a comprehensive review of the key technological foundations that support our research on the AV-pedestrian interaction. Specifically, it focuses on four critical aspects: human behavior data collection methods, human behavior analysis, pedestrian behavior modeling and AV decision-making approaches.

Section 2.1 reviews various methods used to collect vehicle and human behavior data, which serve as an empirical basis for understanding and modeling pedestrian dynamics. Section 2.2 examines existing research on human behavior, with particular emphasis on factors that influence driver and pedestrian decision-making processes. Section 2.3 reviews representative pedestrian behavior models, summarizing key approaches and their capabilities in simulating pedestrian motion and interactions with the AV. Section 2.4 presents an overview of current AV decision-making techniques, ranging from rule-based frameworks to data-driven approaches, and discusses the limitations of existing methods in handling human-aware interaction. Section 2.5 provides a review of relevant work on human-robot interaction.

Drawing on insights from these domains, this chapter summarizes the current state of the art, highlights its limitations, and identifies the research gaps that our work seeks to address in the subsequent chapters.

2.1 Vehicle-Pedestrian Interaction Data

In order to understand complex interaction mechanisms between the vehicle and pedestrian in real-world environments, the availability of high-quality empirical data is essential. As a result, how to effectively collect representative and meaningful data has become a critical question in this research domain.

Recent research has explored various techniques, including questionnaires, video recordings, natural driving studies, and virtual reality technology, to investigate the vehicle-pedestrian interactions.

A questionnaire-based approach was employed in [16], where the survey designed based on the Theory of Planned Behavior to assess pedestrians' intentions to cross the street in particular scenarios and uncover the motivational factors behind these intentions. Study [17] investigated how pedestrians differ in their behavioral responses with potential vehicle-pedestrian collision situations using a questionnaire-based method.

Moreover, Lalam et al. [18] analyzed video data to extract information about vehicle-pedestrian interactions. They then used this data to conduct a multiple linear regression analysis, identifying various factors that impact the severity level of these interactions. Drone footage was employed to gather information on how private cars and taxis interact with pedestrians at unsignalized crosswalks [19]. The study found that most human drivers were willing to yield to pedestrians when sufficient reaction time was available. Vehicle-pedestrian interaction scenarios at three road intersections in Rome, Italy, were recorded using video cameras [20]. Through image processing algorithms, a dataset comprising 270 interaction instances was extracted to facilitate the analysis of interaction behaviors.

Furthermore, a study was conducted on the impact of three factors on safety indicators when a vehicle came across a pedestrian, based on field tests and naturalistic driving data [21]. Natural driving data gathered from real-life situations were also examined to understand how drivers behave when passing pedestrians [22]. Based on a large-scale naturalistic driving dataset, 325 pedestrian crossing scenarios and 23 vehicle non-yielding cases were extracted [23]. The analysis focused on investigating the relationship between vehicle acceleration/deceleration patterns and pedestrian crossing behavior.

Questionnaires are generally considered an easy and straightforward method for collecting interactive data from pedestrians and autonomous vehicles [6]. They enable researchers to access participants' perception, intentions, and attitudes in various traffic scenarios. However, questionnaires may not be ideal for fully and accurately capturing actual road user behavior, due to inherent limitations such as potential subjectivity and recall limitations. For example, when a survey or interview is conducted after an incident or interaction, participants may subjective interpretations or omit important details when answering questions, resulting in potentially unreliable data.

In contrast, video recordings offer the advantage of objectively capturing the actual vehicle–pedestrian interaction as it unfolds. They allow researchers to repeatedly view the recorded scenarios and perform detailed frame-by-frame behavioral analysis, thereby reducing the influence of subjective bias. Compared to questionnaires, video-based methods provide a more faithful representation of observable behaviors. However, video analysis can be time-consuming and labor-intensive, especially when manual annotation is required. Moreover, such methods may lack access to precise kinematic information, such as accurate speed, position, or acceleration, which are crucial for quantitative modeling.

Naturalistic driving data presents a valuable alternative that more accurately reflects real-world vehicle and pedestrian behavior under spontaneous and unconstrained conditions. This type of data is typically collected from on-board sensors and cameras during everyday driving, offering rich, high-fidelity information including speed, acceleration, positional coordinates, and detailed interaction sequences. While naturalistic data provides a more accurate representation of vehicle and pedestrian behavior, obtaining such information through real-world experiments can raise significant safety concerns due to the unpredictable and uncontrollable nature of the process. Researchers cannot anticipate or manipulate when or how risky interactions occur.

In contrast to these techniques, virtual reality (VR) technology enables the construction of virtual worlds where data collection would be both secure and affordable [24]. To gain a deeper understanding of complex pedestrian behavior, VR technology was applied [25] to investigate pedestrian crossing behavior on narrow roadways in residential areas. Researchers also observed the behavior of pedestrians crossing the road at intersections in a virtual reality environment, where an AV was approaching. Their study highlighted the value of virtual reality as an effective tool to analyze pedestrian behavior [26]. Kai Tian et al. at University of Leeds utilized VR experiments to examine how implicit signals from AVs influence pedestrians' crossing decisions [11]. At Queens University Belfast, a virtual reality framework was proposed to provide a safe and cost-effective approach for collecting AV–pedestrian interaction data [27]. A recent study from Delft University of Technology employed a VR setup to investigate AV interactions with multiple pedestrians in shared spaces [28]. This work extended prior research by enabling real-time multi-agent interaction through an innovative VR system. In [29], the authors utilized VR to investigate pedestrian behavior during potential collision events with vehicles. By collecting motion data and analyzing pedestrian walking speeds, they aimed to understand behavioral responses in the presence of perceived risk. Bingbing Nie and colleagues from Tsinghua University developed a VR-based framework to explore vehicle–pedestrian conflicts under emergent traffic conditions [30]. This framework enabled controlled studies of emergency scenarios that would be unsafe or impractical to examine in real-world experiments.

VR technology makes it possible for participants to get fully immersed in controlled experiment scenarios without being exposed to real-world risks, thereby ensuring safety throughout the process. Unlike naturalistic driving data collection, VR can be used to simulate rare or hazardous situations, such as potential collision cases, which are difficult or unsafe to be conducted in real traffic conditions. By simulating realistic traffic environments and dynamic interactions, VR allows researchers to conveniently manipulate key variables such as vehicle position, speed or road layout, while capturing participant responses in real time. Moreover, VR supports the collection of precise behavioral and motion data, including gaze patterns, walking trajectories, and kinematic information, which are essential for analyzing decision-making processes during vehicle–pedestrian interactions. Therefore, VR plays a central role in our study to carry out trials by providing a safe, flexible, and data-rich platform for conducting human-in-the-loop trials in complex urban interaction environments.

2.2 Human Behavior Analysis

2.2.1 Pedestrian Behavior Analysis

Pedestrians considered vulnerable road users, are prone to injury in collisions with vehicles due to the absence of protective equipment. Studying pedestrian behavior is essential for understanding their decision-making process during interaction with vehicles. This understanding holds significant implications for the advancements of AV. If a vehicle is able to interpret a pedestrian's decision-making process, it can better infer their intentions and respond in an appropriate and timely manner. Such anticipation enables the AV to adapt its behavior proactively, whether by slowing down, stopping, or adjusting its trajectory, thereby reducing the probability of traffic accidents and collisions.

Many studies have identified the various factors that influence pedestrian decision-making, including personal attributes, dynamics factors, traffic characteristics, and physical context.

Personal Attributes. Pedestrians' decision-making processes can be influenced by demographics such as age and gender. Risk perception for road crossing varies with gender, with women tending to be more cautious and risk-averse compared to men [31]. Elderly individuals also exhibit a greater likelihood of yielding to approaching vehicles while crossing the street compared to younger adults [32]. Furthermore, psychological traits including sensation seeking and social value orientation proved to impact pedestrians'

decision-making when crossing the road. Kalantari et al. [33] applied a generalized linear mixed-effects model to investigate the impact of various factors including sensation seeking and social value orientation on participants' likelihood to cross the road first. According to [34], pedestrians who have lower levels of sensation seeking tend to be more cautious and may miss opportunities to cross the road compared to those with higher levels of sensation seeking. Besides, pedestrians' walking trajectories can also be influenced by their attentiveness [35]. When compared to undistracted walkers, distracted pedestrians, such as those using mobile phones, frequently alter their walking direction and travel more slowly.

Dynamics Factors. Gap acceptance, usually measured in time, is considered a critical dynamic feature for evaluating pedestrian safety during road crossings. This acceptance is connected to both the vehicle's speed and the longitudinal distance between the vehicle and pedestrians, often represented as Time-to-Arrival (TTA) [36]. Pedestrians typically accept gaps for crossing that range from 3 to 7 seconds. If the gap is less than 3 seconds, pedestrians generally opt not to cross [37], while gaps exceeding 7 seconds prompt pedestrians to cross without hesitation [38]. Alterations in TTA have also been regarded as effective non-verbal cues between drivers and pedestrians during communication [39]. A study analyzing various scenarios involving vehicle-pedestrian interactions through naturalistic driving research has shown that vehicle speed and distance are significant factors influencing these interactions [40]. A more recent study by Tian and Markkul [41] investigated that pedestrians are more likely to cross the street when vehicles drive at higher speeds, leading to a higher proportion of unsafe crossings. In [11], it was discovered that the deceleration parameter can serve as a form of communication between humans and autonomous vehicles. Findings from [42] suggest that pedestrians tend to lose patience with longer waiting time and are more inclined to accept smaller gap durations for crossing.

Traffic Characteristics. Research highlights the significant impact of traffic volume on pedestrian decision-making. Studies, such as [43], indicate that higher traffic volumes reduce the likelihood of people choosing to cross roadways. The traffic volume and pedestrian waiting time are positively correlated [44]. Vehicle type and size also have an impact on pedestrian crossing behavior. Pedestrians exhibit varying acceptable waiting time based on different vehicle types. For instance, they frequently cross in front of passenger cars at a high speed [45]. In developing countries, heavy vehicles, like trucks, tend to have a lower yielding rate compared to light vehicles [46] due to their increased risk associated with longer braking distances [47]. Typically, pedestrians exercise more caution when crossing in front of larger vehicles [48].

Physical Context. Pedestrian behavior while crossing roads is shaped by multiple physical context factors, including road characteristics, weather conditions, and the time of day. Pedestrians tend to increase their crossing speed on wider roads with more lanes. However, they exhibit more caution on wide roads and are inclined to accept smaller gaps on narrower roads [49]. In unsignalized crosswalks, pedestrians were observed to engage in more frequent traffic checking, move at reduced speeds, and encounter a higher number of vehicle-pedestrian conflicts [50]. Additionally, inclement weather conditions prompt pedestrians to adopt a more conservative approach when crossing roads [51]. An observable trend is the increased number of individuals crossing roads during rush hour [45].

While pedestrian personal information and mental state indeed affect their decisions, collecting such information in real-world scenarios remains a significant challenge for AVs. Unlike observable physical features, individual states such as age or emotion are not directly accessible through onboard sensors. This limitation makes it impractical to rely on such data for real-time inference. Moreover, it has been proposed that personal traits have been implicitly incorporated into motion cues [39]. For example, pedestrian walking speed can be influenced by factors such as age, gender, emotions, and social status. Elderly pedestrians may walk more slowly due to physical limitations, while individuals experiencing uncertainty or hesitation may exhibit irregular trajectories or frequent stops. Furthermore, a road crossing experiment conducted in [52] involving twenty participants highlighted the significance of motion cues influencing pedestrians' behavior. Similarly, the findings in [53] demonstrate that pedestrians and drivers rely more on kinematic cues to make decisions. Kinematic information has also been found to be more important than psychological characteristics like sensory seeking and social value orientation when deciding who crosses unmarked intersections first [33].

To better focus on understanding the underlying mechanism of vehicle-pedestrian interaction, this thesis temporarily sets aside the influence of contextual factors such as traffic characteristics and physical context. While such factors undoubtedly play a role in real-world behavior, their complexity and variability introduce considerable noise that can obscure the core dynamics of interaction. Instead, our thesis concentrates on the kinematic parameters that affect how pedestrians behave as it represents a more accessible and practical source of data.

2.2.2 Human Driver Behavior Analysis

Understanding human driver behavior is fundamental to the development of safe and intelligent autonomous driving systems, especially in complex interactive scenarios involving pedestrians and other road users. Similar to pedestrian behavior, driver behavior is also influenced by a wide range of factors, which can be categorized into individual, kinematic and environmental dimensions [54].

Individual characteristics, such as age, gender, driving experience, and personality, can shape a driver's perception of risk and decision-making tendencies [55]. Kinematic factors, such as vehicle speed, pedestrian walking speed and inter-agent distance, play a crucial role in shaping driver behavior during interactions with pedestrians. Study [40], which examined real-world interaction cases of using Logistic regression analysis, identified vehicle speed and the distance to pedestrians as key predictors of drivers' behavioral responses. Moreover, the study [21] employed naturalistic driving data to investigate drivers' overtaking behavior on rural roads. The results showed that drivers tend to maintain smaller lateral gaps and adopt lower overtaking speeds when pedestrians are walking in the opposite direction, positioned near the edge of the lane, or when there are oncoming vehicles. Environmental factors, including road layout, visibility, and weather conditions, can have a direct impact on driving behavior [56]. For example, the authors in [57] analyzed real-world driving data to quantify the impact of environmental conditions on driver behavior. Their findings revealed that rainy weather significantly alter driving patterns, as drivers tend to adapt their actions to maintain safety under adverse weather.

While factors such as age, driving experience, and personality contribute to variability in individual driving styles, this thesis aims to characterize a representative behavioral model rather than account for person-specific distinctions. Therefore, the focus is placed on understanding how kinematic variables influence driver behavior during interactions with pedestrians. These variables are directly observable and provide a foundation for developing reliable, real-time decision-making frameworks for AVs in complex urban settings, particularly at unsignalized intersections and within shared spaces where such interactions are frequent.

Beyond identifying the factors that influence human drivers' decisions, cognition models were developed to describe and predict driver behavior. One widely studied approach is the risk field model [58], which represents the surrounding environment as a dynamic field of potential risks, with drivers assumed to select actions that minimize their perceived risk. Kolekar et al. [59] introduced a driver risk field model, which represented the driver's perceived risk as a subjective probability field combined with potential consequences, and

demonstrated that this framework can reproduce human-like driving behaviors across diverse scenarios. In [60], a driver risk field model was proposed to simulate car-following behavior, where driver strategies are derived from the evaluation of potential risks in the surrounding environment. Another category of driver models is the perception-based model, which employs perceptual cues (e.g., vehicle speed, distance, gap acceptance, and visual looming.) to describe drivers' internal decision processes [61]. For instance, study [62] proposed an evidence accumulation framework based on visual looming to account for braking decisions in rear-end collision scenarios. Similarly, study [63] introduced an evidence accumulation model for gap acceptance in left-turn situations, demonstrating that dynamic evidence accumulation underlies human drivers' decisions to accept or reject a turning gap. These models provide valuable insights into human driver behavior and offer strong interpretability. However, they cannot be directly applied as AV decision models, since they are primarily designed to explain human behavior rather than to guarantee safe navigation. Moreover, AV-pedestrian interactions are characterized by high levels of uncertainty and unpredictability, which makes it difficult for such models to capture the full variability of real-world behavior. In contrast, data-driven approaches [64], which leverage large amounts of driving data to learn behavioral patterns, are well suited to handle dynamic and uncertain environments, but often suffer from a lack of interpretability. Another important class of cognition models is Bayesian inference models, which describe how humans infer others' intentions through belief updates based on observed behavior. Study [65] developed a Bayesian decision-making model for AV lane-changing scenarios, where the lane-change decision is continually updated through Bayesian inference and executed once the inferred decision variable exceeds a predefined threshold. Likewise, a Bayesian driver agent model was developed for AV system to interpret the surrounding environment and make decisions in a manner inspired by the human cognitive psychology [66]. While these Bayesian models are more interpretable through explicit inference and belief update modules, they do not account for risk perception or translate perceived risk into behavioral responses. Therefore, this study aims to integrate risk-based and cognitive modeling principles into data-driven frameworks, enabling models that can both handle uncertainty in complex traffic scenarios and provide enhanced interpretability.

2.3 Pedestrian Behavior Modeling

Pedestrians are dynamic agents with complex and often unpredictable trajectories, requiring AVs to continuously account for their movements during decision-making processes. Consequently, pedestrian behavior modeling plays a critical role in the development of AV decision-making algorithms in human-involved environments. By capturing the behavioral characteristics and decision-making patterns of pedestrians, AVs can better

interpret and adapt to human actions, enabling more timely and appropriate responses during interactions. Accurate pedestrian models serve as the foundation for realistic interaction simulation, providing meaningful feedback that allows AV policies to be developed and evaluated in environments that closely reflect real-world complexity. This contributes to more robust and generalizable decision-making strategies. In recent years, a wide range of pedestrian modeling approaches have been proposed to facilitate the simulation and analysis of human-vehicle interactions. This section presents a review of representative pedestrian behavior modeling approaches, with a focus on their assumptions, modeling techniques, and applications in supporting AV decision-making in interactive contexts.

Early approaches often rely on simplified pedestrian models with ideal behavioral assumptions that significantly deviated from real-world scenarios. For instance, [12] modeled pedestrians as moving at a constant velocity and fixed direction without reacting to surrounding dynamics. The Cellular Automaton (CA) model, first introduced in [67], discretized the current space into a grid of cells, where each cell can accommodate at most one pedestrian. Pedestrian movements were governed by simple, rule-based transitions between adjacent cells, considering factors such as desired direction, occupancy of neighboring cells, and local interactions. This model simulates pedestrian dynamics in discrete situations, lacking the ability to represent the continuous motion.

Pedestrian reactive models, such as the Velocity Obstacle (VO) model, the Optimal Reciprocal Collision Avoidance (ORCA) or Social Force Model (SFM), incorporate agent interactions and collision avoidance capabilities to simulate pedestrian motion in dynamic environments. The VO model [68] defined a set of forbidden velocities that would lead to a future collision with moving obstacles. By avoiding these velocity regions, an agent could select a motion that approximates its desired speed toward the destination. Despite its simplicity, intuitiveness, and computational efficiency, the VO model typically assumes that other agents maintain a constant velocity over time without considering their potential reactions or avoidance behaviors. Compared to the VO model, ORCA [69], [70] formulated a reciprocal collision avoidance strategy, where each agent was responsible for adjusting its velocity to avoid collisions cooperatively with others. The ORCA model performs well and captures reactive behavior among agents, but it still depends on assumptions that other agents will reciprocate collision avoidance efforts in a cooperative manner and prioritizes obstacle avoidance over goal-directed intentions. Another widely adopted approach is the SFM, originally proposed by Helbing and Molnar [71]. In this model, pedestrians were treated as particles influenced by a set of social forces that guide their motion [72], [73]. These includes a driving force that propels the pedestrian toward a desired destination and repulsive forces to maintain distance from other pedestrians and obstacles. The resulting motion emerged from the continuous summation of these forces, allowing for simulations of pedestrian behavior in various environments. However, the

SFM primarily focuses on spatial proximity when modeling interactions, assuming that pedestrian behavior is solely influenced by distance-based factors. In reality, pedestrian decision-making is affected a broader range of factors beyond physical distance, such as relative velocity.

Data-driven methods learn pedestrian behavior patterns directly from empirical data without relying on hand-crafted interaction rules or pre-defined force models. These methods leverage large-scale trajectory datasets to capture complex and context-dependent motion dynamics that may be difficult to express analytically. For instance, recurrent neural networks (RNNs) [74] and their variants such as Long Short-Term Memory (LSTM) [75], [76] and Gated Recurrent Unit (GRU) [77], [78] networks have been widely applied to predict pedestrian trajectories by modeling temporal dependencies in movement sequences. Recent studies [79], [80] further incorporate social context by encoding surrounding agents' behaviors as inputs to the network, enabling the model to account for implicit interaction patterns among pedestrians. Transformer-based architectures [81], [82] have also been explored to improve long-range dependency modeling and better capture subtle behavioral cues.

These data-driven methods provide greater flexibility in modeling pedestrian behavior, making them well-suited for capturing diverse motion patterns and decision-making dynamics. However, their performance heavily relies on the quality and diversity of the training data. In addition, these models often operate as black boxes, lacking interpretability and transparency in how behavioral predictions are generated. Currently, most data-driven models are primarily integrated into AV systems as trajectory predictors, providing forecasts of pedestrian movements to inform AV decision-making, rather than serving as independent pedestrian agents in interactive simulations. Although a data-driven pedestrian predictor was employed as a pedestrian model for AV-pedestrian interaction simulations in [83], it was trained solely on pedestrian trajectory datasets and thus learned pedestrian-pedestrian interactions. Due to the absence of vehicle information during training, the model failed to account for pedestrian reactions to AV maneuvers.

In addition to general pedestrian motion models, some studies have specifically focused on modeling pedestrian crossing decisions. A common approach treats crossing behavior as a gap acceptance process, where the pedestrian decides whether to cross based on the perceived gap between themselves and approaching vehicles. For instance, [84] proposed a crossing decision model based on the distance and velocity of oncoming vehicles, while [85] adopted the minimum Time-to-Collision (TTC) as a threshold to determine crossing safety. However, such models often lack a solid psychological foundation and tend to oversimplify human decision-making by relying solely on physical variables. To address this limitation, researchers have introduced psychophysically inspired models that incor-

porate perceptual cues. Domeyer et al. [61] introduced a pedestrian model that relied on visual angle, its derivative (looming), time-to-arrival, and bearing angle to characterize how pedestrians perceived approaching vehicles and made crossing decisions, with model parameters calibrated using naturalistic driving data. [41] presented a gap acceptance model that employed visual looming information together with binary logistic regression to infer crossing intentions. More recently, [86] developed a crossing decision model driven by perceptual variables such as the rate of change of visual angle and the ratio τ (visual angle to its rate of change), and showed through simulations that it can reproduce pedestrian crossing patterns across different vehicle kinematics. The evidence accumulation model [87], inspired by cognitive psychology, assumed that crossing decisions resulted from a gradual accumulation of perceptual evidence over time, with a decision triggered once an internal threshold was reached. Despite their contributions, most of these models simplify pedestrian behavior to discrete decisions, such as whether or when to cross, rather than predicting continuous motion trajectories or capturing dynamic, real-time interactions. Moreover, they are typically designed for structured traffic scenarios (e.g., marked crosswalks or unsignalized intersections). Their applicability to more complex or unstructured environments remains limited, where pedestrian behavior is inherently continuous, and not constrained to binary crossing choices.

In contrast to the aforementioned pedestrian models, study [13] used recorded pedestrian trajectories, which inherently captured human-human interactions, to train AV policies. While these trajectories provide realistic motion patterns, they remain static during simulation and do not adapt to the AV's behavior in real time.

In summary, existing pedestrian models have laid the groundwork for simulating human behavior in traffic environments. However, most remain limited in their ability to support real-time, bidirectional interactions with autonomous vehicles. In contrast, the pedestrian models proposed in this study are specifically designed to enable dynamic and continuous interaction with the AV during simulations. More importantly, they incorporate cognitive modeling to reflect how pedestrians perceive, reason about, and adapt to surrounding agents. These models not only facilitate more realistic and responsive interactions but also enhance interpretability in understanding how specific behaviors emerge.

2.4 AV Decision-Making Approach

Numerous decision-making algorithms have been proposed to enable AV to navigate complex urban environments and interact safely with surrounding agents. These algorithms are critical for translating perception inputs into executable actions, ensuring that the AV can operate efficiently and safely under diverse scenarios. Existing AV decision-making approaches can be classified into three categories: rule-based methods, which rely on manually defined heuristics and traffic norms; utility-based methods, which optimize predefined cost or reward functions to select the most beneficial action; and learning-based methods, which employ machine learning, particularly deep reinforcement learning, to learn policies directly from data or simulation. This section delves into a detailed description of these approaches.

2.4.1 Rule-Based Method

Rule-based methods rely on a database of predefined rules derived from traffic legislation, expert driving experience, and automotive knowledge. The AV continuously gathers real-time information regarding its surrounding environment and matches it with the empirical knowledge of the behavior rule database, thus making a reasonable decision for the next moment [88]. A typical representative of this approach is the finite state machine (FSM), a mathematical model consisting of a set of states and transfer relationships among them. Figure 2.1 illustrates a simple finite state machine model of AV lane changing [89]. According to the logical structural relationship between states, FSM models are classified into series [90], parallel [91], and hybrid types [92], which have been widely applied in AVs [89]. For example, Stanford University's AV, known as Junior, utilized a parallel finite state machine model for behavioral decision-making during the 2007 Defense Advanced Research Projects Agency (DARPA) competition. This FSM model involved transitions between 13 states, such as driving forward and passing through the intersection [93]. Fuzzy logic represents an extension of traditional rule-based approach. Although it still remains rule-driven, fuzzy logic enables more flexible reasoning by allowing partial truth values and soft boundaries in rule definitions. For example, study [94] proposed a hierarchical fuzzy rule-based system to determine the appropriate passing speed when navigating intersections in the presence of human-driven vehicle. In addition, model predictive control (MPC) has been widely adopted in autonomous driving as a rule-based method. Its core idea is to use a predictive model of vehicle dynamics to forecast the evolution of the system

over a finite horizon and then solve an optimization problem that minimizes a predefined cost function subject to safety and physical constraints [95]. Typical applications include nonlinear MPC for trajectory tracking and obstacle avoidance [96], and stochastic MPC for lane changing [97].

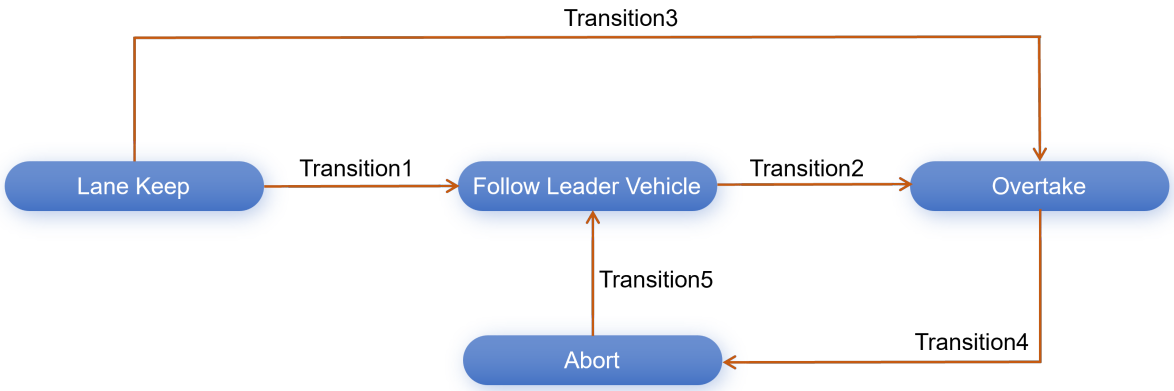


Figure 2.1: A simple finite state machine model of AV overtaking maneuver.

Rule-based decision-making methods are simple, easy to implement, highly interpretable, and perform better in simple scenarios. However, these methods are less successful in real-world urban driving scenarios [98] due to dynamic and complex road conditions. This is because it is difficult for a finite set of rules to cover all possible scenarios, especially in complicated and uncertain traffic environments, where these rules may have limitations. For example, when a vehicle navigates in a busy urban intersection, a rule-based approach may struggle to handle the changing traffic circumstances and uncertain human activities, making it challenging to guarantee both the safety and efficiency of traveling simultaneously.

2.4.2 Learning-Based Method

Learning-based methods have emerged as powerful tools for AV decision-making, enabling systems to learn optimal policies directly from data and interaction with the environment. Unlike rule-based approaches that rely on manually designed rules, learning-based techniques offer greater flexibility and adaptability, especially in dynamic and uncertain traffic scenarios. This section reviews the learning-based methods, which can be categorized into two major groups: imitation learning, and reinforcement learning.

2.4.2.1 Imitation Learning

Imitation learning (IL), a supervised machine learning paradigm, learns optimal behavior by observing and mimicking expert demonstrations [99]. In the field of AV, imitation learning typically involves collecting a dataset of human driver behaviors, including steering, acceleration, and braking in various traffic scenarios. This collected operational data is then used to train the self-driving system, enabling it to mimic and learn from these expert behaviors for safe and stable driving [100]. Unlike reinforcement learning, it does not rely on trial-and-error exploration to optimize performance; instead, it assumes that expert behavior provides a near-optimal solution.

Behavioral cloning (BC) and inverse reinforcement learning (IRL) are typical representative methods within imitation learning. Behavioral cloning directly maps sensor inputs to control actions by learning from expert demonstrations [101]. One of the earliest implementations of behavioral cloning in AV was proposed by Pomerleau et al. [102], who employed a simple three-layer neural network to map raw images captured from a camera to steering commands, laying the foundation for later end-to-end driving approaches. A well-known example is the work of Bojarski et al. [103] at NVIDIA, where a convolutional neural network (CNN) was trained to learn a direct mapping from raw front-facing camera images to steering control outputs. This end-to-end system allowed the AV to drive on both regular roads and highways without requiring intermediate modules such as lane detection, path planning, or control logic.

Later extensions introduced richer inputs and temporal reasoning. Hecker et al. [104] developed a system equipped with eight cameras to capture a 360-degree view around the vehicle. The visual inputs from four selected cameras were processed through a network consisting of multiple CNN and LSTM modules, and a final fully-connected layer fused information from all camera views along with map data to predict future speed and steering. This significantly improved performance by leveraging richer contextual information. Similarly, the authors [105] combined CNNs with recurrent layers to train an AV in urban environments. This study found that incorporating temporal information through the recurrent structure significantly improved the model's performance. Another study [106] proposed a CNN-LSTM model that used the past 1.5 seconds of image and state history to predict a 3-second collision-free trajectory, including both speed and position. Recent advances have also explored integrating Transformer encoders with CNNs to improve spatial-temporal modeling for predicting steering and velocity [107].

To enhance task-specific control, Codevilla et al. [108] proposed Conditional Imitation Learning (CIL), which introduces high-level navigation commands (e.g., turn left, go straight, turn right and continue) as additional conditional inputs, as shown in Fig. 2.2. The model employs a branched architecture where each branch corresponds to a specific command, and the appropriate branch is activated at inference time, allowing the system to adapt its behavior to different driving tasks.

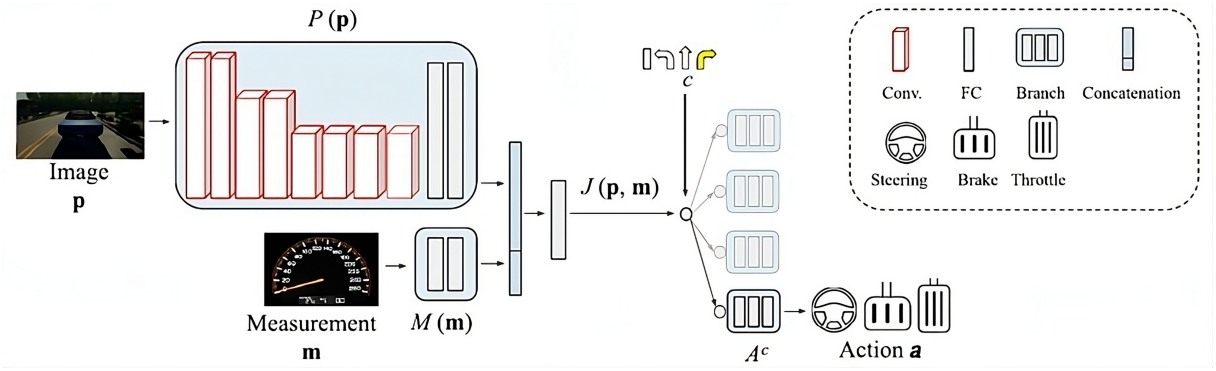


Figure 2.2: CIL architecture (from [108]).

Although BC has been successfully applied to certain driving situations due to its simple implementation and training efficiency, it is crucial to acknowledge several limitations. A substantial volume of supervised data is required to be provided for learning [107], which is costly and time-consuming to collect. Another critical limitation is its strong dependence on the quality and coverage of expert demonstrations [109]. If specific scenarios are not available in the training data, the model may fail to generalize and perform reliably in unfamiliar situations. Moreover, BC suffers from the covariate shift problem [110]: once the model deviates from the expert trajectory during inference, the resulting errors may accumulate over time and lead to significant failures.

In contrast, IRL aims to deduce the potential reward function behind the human driver's behavior [111]. By observing a set of expert behaviors, IRL algorithms estimate a reward function under which the demonstrated behavior would be considered optimal. This inferred reward is then used to train a policy through reinforcement learning, allowing the agent to adapt its behavior even in novel or slightly different situations.

The study [112] was the first to formally define the IRL problem. It proposed a linear programming-based approach to recover a reward function from expert demonstrations, ensuring that the expert's behavior appears optimal under the inferred reward. This foundational work established the theoretical basis for IRL by shifting the focus from learning direct policies to uncovering the underlying motivations that guide expert actions. Ziebart et al. [113] proposed Maximum Entropy IRL, to address the ambiguity and overconfidence often observed in traditional IRL methods when modeling driving behaviors. By introducing the maximum entropy principle, this method inferred a reward

function that explained expert behavior while preserving the stochasticity of real-world decision-making. This allows the model to better capture the variability and uncertainty present in expert demonstrations. Building on this foundation, Wulfmeier et al. [114] extended the Maximum Entropy IRL framework by integrating deep neural networks to learn reward functions directly from high-dimensional sensory inputs . This approach was subsequently applied to complex path planning tasks in urban roads [115], [116].

Ho and Ermon later proposed Generative Adversarial Imitation Learning (GAIL) [117], which combines the principles of IRL and Generative Adversarial Network (GAN). Rather than explicitly learning a reward function, GAIL trained a discriminator to distinguish expert behaviors from those generated by the learned policy, enabling the policy to imitate expert behavior directly. By leveraging the adversarial learning paradigm, GAIL effectively mitigated the issue of state distribution drift arising from limited demonstration data [118]. Based on GAIL, Li et al. proposed the InfoGAIL framework [119], which introduces latent variables to capture and disentangle distinct driving styles. This framework allows for the explicit modeling of diverse behavior patterns within expert demonstrations.

IRL provides strong generalization capabilities and high interpretability by recovering reward functions that reflect the underlying motivations of expert behavior. However, it is computationally intensive due to its reliance on nested reinforcement learning, which becomes especially costly in high-dimensional and continuous state spaces. Moreover, similar to behavioral cloning, IRL is highly sensitive to the quality of expert demonstrations. Biased or suboptimal behavior can lead to incorrect reward inference and ultimately degraded policy performance.

2.4.2.2 Reinforcement Learning

Reinforcement learning (RL) typically employs the Markov Decision Process (MDP) framework [120] to model the interaction between an agent and its environment. In this framework, the agent observes the current state, select an action based on a policy, and transitions to a new state while receiving a corresponding reward, as shown in Fig. 2.3. Through this continuous trial-and-error process, the agent seeks to learn an optimal policy that maximizes the expected cumulative reward over time. In the domain of autonomous driving, reinforcement learning provides a promising framework for learning complex decision-making behaviors in dynamic and uncertain environments. The AV acts as the agent, while the environment includes surrounding pedestrians, other vehicles and road infrastructure. The agent makes sequential decisions based on information about the surrounding environment from the vehicle's sensors as well as the rewards from the previous action. Reinforcement learning has received significant attention for its application in the

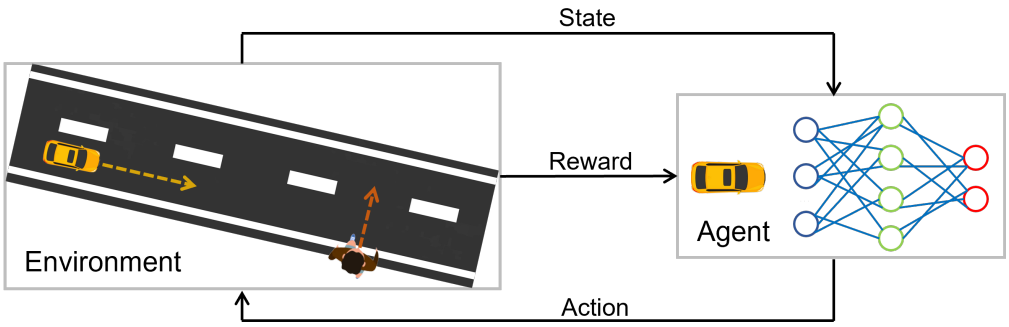


Figure 2.3: Single-agent reinforcement learning framework for autonomous driving

autonomous driving decision-making. Existing methods can be classified into value-based, policy-based, and actor-critic approaches. Value-based methods focus on estimating the expected return of taking an action in a given state and selecting actions greedily based on these value estimates. A Q-learning algorithm was proposed in [121] to enable an AV to maneuver through roundabout scenarios. This method was verified through simulations in the CARLA simulation environment, demonstrating that the AV equipped with this algorithm could navigate roundabouts safely and effectively. A deep Q-learning (DQN) model with a discretized action space was employed in [122] to train an AV to perform lane-keeping while maintaining a safe distance from leading vehicles in a simulated urban driving environment. However, the original DQN suffers from an overestimation bias in Q-values [123], caused by using the same network for both action selection and evaluation. To mitigate this issue, Double Deep Q-Learning (DDQN) was introduced, which decoupled these two processes by using the online network to select the action and the target network to evaluate its value. Experimental results in driving simulators have shown that DDQN outperforms DQN in terms of learning stability and decision-making effectiveness [124]. While effective in discrete action spaces, value-based methods often struggle with high-dimensional or continuous action domains due to the difficulty of accurately estimating the value function over large or infinite action sets. In such settings, discretizing continuous actions can be computationally expensive or even intractable, and often leads to suboptimal performance, especially in tasks with continuous action spaces, such as autonomous driving.

Different from value-based methods, policy-based methods directly optimize the parameterized policy without relying on value function estimation, which are better suitable for environments with continuous action spaces. The REINFORCE algorithm, introduced by Williams [125], is one of the earliest Monte Carlo policy gradient methods. It directly updates a policy by estimating the gradient of the expected return from sampled trajectories. Despite its simplicity, REINFORCE suffers from high variance and poor sample efficiency [126], as it relies solely on Monte Carlo returns and does not incorpor-

ate any form of value function estimation. These limitations motivated the development of more advanced policy optimization techniques, particularly those based on the actor-critic framework, which introduces a learned value function to serve as a baseline, thereby reducing variance and improving the stability of policy updates.

The actor-critic framework combines the strengths of both value-based and policy-based approaches [127]. In this architecture, the actor is responsible for learning a policy, while the critic estimates a value function to evaluate the quality of the actor's decisions and provide informative feedback to guide policy updates. By integrating policy learning and value estimation, actor-critic methods achieve more stable and sample-efficient training, making them particularly well-suited for autonomous driving. For example, the study [128] first implemented the Deep Deterministic Policy Gradients (DDPG) algorithm in a lane following scenario. As a model-free, off-policy actor-critic method, DDPG is specifically designed for learning deterministic policies in continuous action spaces. It employed a single critic network to estimate Q values, guiding the actor network in producing continuous actions including vehicle speed and steering angle. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm was employed in [129] for training an AV for navigating T-shaped intersections. As an improvement over DDPG, TD3 incorporated two independent critic networks and uses the minimum of two Q values estimates to mitigate overestimation bias. This adjustment thus enhanced learning stability and training performance. The Soft Actor-Critic (SAC) algorithm was adopted to train an AV for lane changing on highways [130]. SAC was an off-policy actor-critic algorithm that also utilized two critic networks to improve the reliability of value estimates. This approach also introduced maximum entropy framework, which encouraged the agent to explore a diverse range of behaviors by maximizing both the expected return and the policy entropy. This characteristic enabled the AV to learn flexible lane-changing strategies. The authors in [131] utilized the Proximal Policy Optimization (PPO) algorithm, a widely used on policy method, to address the AV on-ramp merging task. PPO introduced a clipped surrogate objective to restrict policy updates within a trust region, which effectively improved training stability and prevented performance collapse due to overly aggressive updates. Maramotti et al. [132] used Asynchronous Advantage Actor-Critic (A3C) algorithm to develop an end-to-end driving policy directly from raw image inputs. By distributing multiple agents across separate environment instances, the framework enabled asynchronous policy and value updates to a shared global model. This design not only accelerated the learning process but also mitigated sample correlation, thereby removing the reliance on experience replay buffer.

In complex urban driving environments, conventional RL architectures that often based on simple multilayer perceptrons (MLPs), may struggle to capture the intricate spatial and temporal patterns inherent to the task. To address this limitation, recent research has integrated advanced representation learning modules into RL frameworks, such as Graph Neural Networks (GNNs) for structured interaction modeling, RNNs for temporal sequence processing, and attention-based models like Transformers for capturing long-range contextual information.

For example, the authors in [133] augmented the DQN framework with Graph Convolutional Network (GCN) to explicitly model interactions among multiple agents during highway ramp merging. By representing the traffic scene as a graph, where nodes denote vehicles and edges encode their spatial relationships, this model can better reason about complex inter-agent dynamics and make more informed decisions. Similarly, [134] integrated both GCN and Graph Attention Network (GAT) into the PPO framework to capture spatial relationships among surrounding agents more effectively. Given that decision-making in autonomous driving is inherently sequential, RNNs such as LSTM and GRU have been adopted to model temporal dependencies in decision-making. LSTM was utilized in [135] to support lane-changing behavior on highways, while GRU was applied in [136] for navigating complex urban intersections. These recurrent structures allow the policy network to retain and leverage past observations, leading to more temporally consistent driving behavior. In addition, the Transformer architecture was incorporated into the SAC framework in [137] to encode high-level scene representations and extract structured contextual cues, thereby enhancing the policy's decision-making capability. Furthermore, [138] proposed an attention-enhanced DQN for highway driving scenarios, where the attention mechanism enables the agent to prioritize critical environmental features, such as nearby vehicles or hazardous regions, thus improving situational awareness and decision-making accuracy.

Uncertainty quantification has emerged as an important topic in learning-based approaches for autonomous driving. Two types of uncertainty are considered: aleatoric and epistemic uncertainty [139]. Aleatoric uncertainty originates from the inherent randomness of the environment or data, such as sudden changes in pedestrian motion or measurement noise from sensors [140]. Epistemic uncertainty, in contrast, arises from the limited knowledge or expressiveness of the model, which leads to poor performance in situations that were not sufficiently represented in the training data [140]. To quantify these uncertainties, a variety of methods have been developed, including Bayesian neural networks [141], Monte Carlo dropout [142], deep ensembles [143], and distributional reinforcement learning [144]. These techniques allow learning-based models to generate predictions alongside confidence estimates, which is critical for ensuring safety and more reliable autonomous driving decision-making systems.

In contrast to the aforementioned single-agent RL approaches, where only the AV is trained as an agent and other road users are modeled with fixed trajectories or rule-based behavior models, MARL treats all interactive entities, such as surrounding vehicles and pedestrians, as learning agents, as shown in Fig. 2.4. In this framework, each agent simultaneously learns its own policy while interacting with others in a shared environment. Unlike single-agent RL, where an agent often focuses solely on maximizing its own reward, potentially leading to locally optimal behaviors or even degraded global performance, MARL enables the optimization of system-level performance or the balancing of individual objectives among agents.

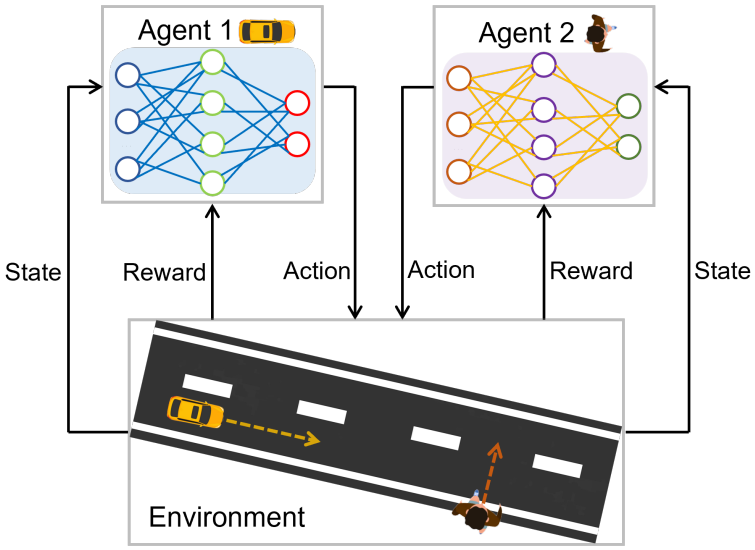


Figure 2.4: Multi-agent reinforcement learning framework for autonomous driving

To support efficient coordination and learning among agents in MARL, different training and execution paradigms have been explored. Centralized Training Centralized Execution (CTCE) [145] provides full observability during both training and execution but is often impractical due to the unrealistic assumption of global information availability during execution. On the other hand, Decentralized Training Decentralized Execution (DTDE) [146] fully reflects real-world constraints by relying solely on local observations, yet it often leads to poor sample efficiency and limited coordination. In contrast, Centralized Training Decentralized Execution (CTDE) [147] framework has emerged as the dominant paradigm in MARL. Under this framework, agents are trained using global information, while during execution, each agent selects actions based solely on its own local observations, as shown in Figure 2.5. This approach balances training efficiency with deployment feasibility.

Several representative algorithms are developed within the CTDE framework. For instance, the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm was adopted to model interactions between an AV and a human-driven vehicle [148]. During training, a centralized critic leveraged the full joint state and action information of all agents to better capture the coupled dynamics between them. The Multi-Agent Proximal Policy Optimization (MAPPO) extended the single-agent PPO algorithm to multi-agent settings by introducing a centralized value function [149]. MAPPO enabled each agent to learn a dedicated policy while relying on centralized value estimation. The Multi-Agent Soft Actor-Critic (MASAC) built on the SAC architecture with a shared entropy-regularized critic, promoting diverse yet coordinated behavior to enhance highway traffic throughput [150]. By allowing agents to adapt to one another, multi-agent reinforcement learning (MARL) facilitates the emergence of more socially compliant and cooperative behaviors. However, despite its advantages, MARL still faces several key challenges. One major issue is the non-stationarity of the environment [151], as each agent's learning process continuously alters the dynamics perceived by others. Additionally, the explosion in the joint state-action space further complicates training and hampers scalability, especially in complex real-world driving scenarios.

As this thesis mainly focuses on AV–pedestrian interactions, recent advances in learning-based methods to AV decision-making in urban settings involving pedestrians are subsequently examined. Study [152] employed a maximum entropy IRL method to recover the underlying reward function from real-world driving data involving vehicle–pedestrian crossing events. This demonstration-driven approach enabled the construction of an interpretable and data-driven model of AV behavior in human-involved environments. A risk-aware RL framework was proposed in [153], in which a risk prediction module was integrated into the decision-making process. This integration allowed the AV to adopt safer strategies in pedestrian interactions. Although the AV's action space was discretized into a set of predefined maneuvers, the model exhibited improved robustness under uncertainty. In [15], a DQN algorithm was used to train the AV policy in scenarios involving distracted pedestrians. Pedestrian behavior was defined by a handcrafted rule-based model, while the AV learned to adapt its responses to balance safety and efficiency. A hybrid decision-making architecture was developed in [154], combining rule-based logic and RL. The AV determined its final action by evaluating the confidence of each policy and selecting the one deemed more reliable, thus enhancing safety in uncertain pedestrian interactions. The SAC algorithm was extended by incorporating Social Value Orientation (SVO), a concept from psychology that quantified an agent's preference between self-interest and prosocial behavior [155]. This augmentation allowed the AV to adjust its behavior based on the social context, with pedestrian behavior simulated using the Social Force Model. To model mutual adaptation between AVs and pedestrians, study [156] formulated the interaction as a MARL problem using a DQN-based framework, where both agents were treated as

learning agents. Similarly, [157] employed the MADDPG algorithm, using a CTDE setup to jointly train the AV and pedestrian. This approach allowed the AV to develop cooperative crossing behaviors by accounting for the evolving pedestrian strategy. While most of the aforementioned works focused on single-pedestrian scenarios, [13] addressed the more complex case of AV interactions with multiple pedestrians in urban shared spaces. Real pedestrian trajectory data were used to simulate the environment, and the AV policy was trained via the PPO algorithm with a reward function designed to ensure both safety and efficiency in dense crowds.

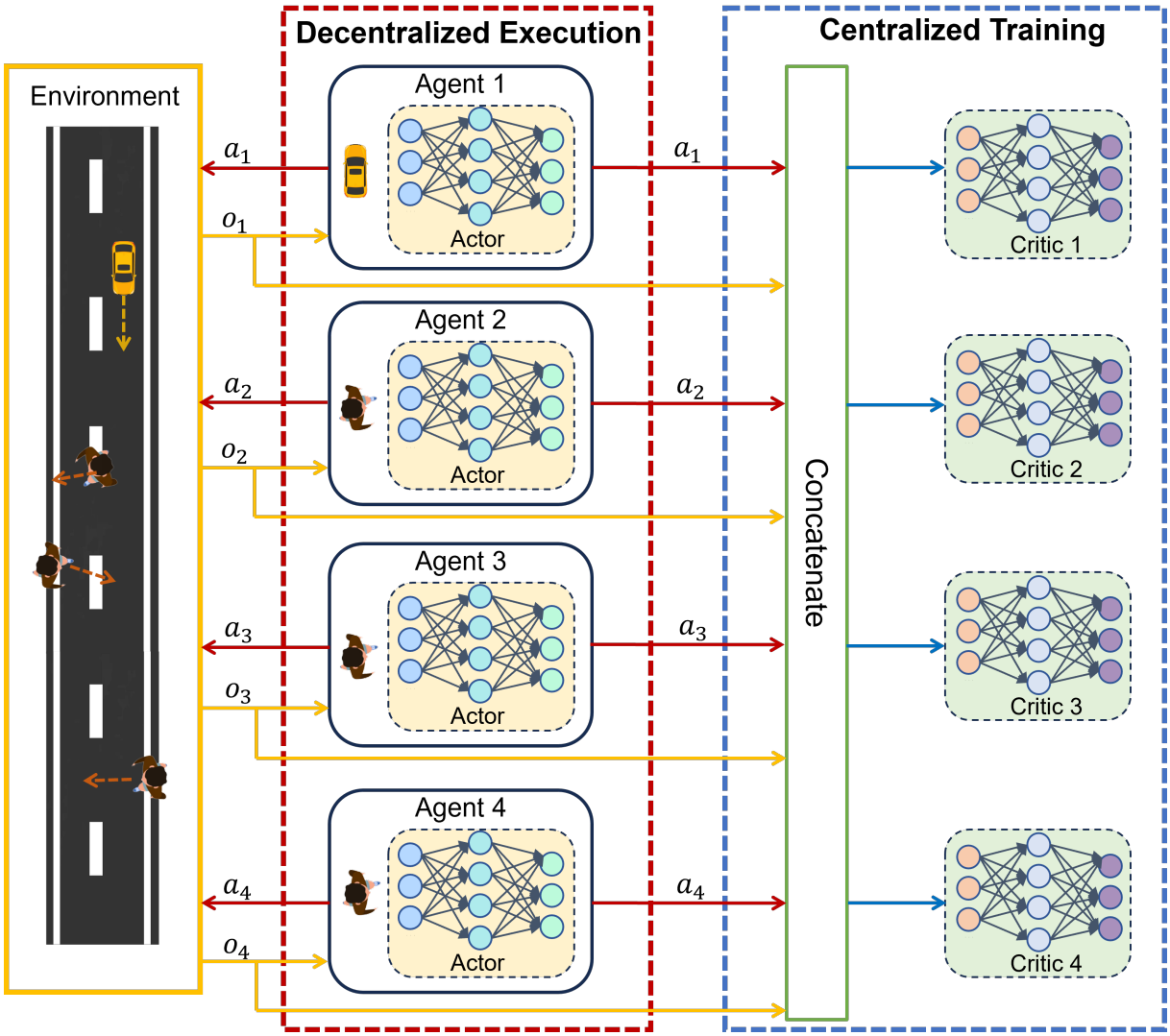


Figure 2.5: An illustrative diagram of the CTDE framework in MARL.

While learning-based approaches can effectively model how an agent interacts with dynamic and complex environments to make decisions accordingly, safety requirements demand absolute accuracy of the AV in real-world situations, which contradicts the ”trial-and-error” of the training process [98]. In addition, deep reinforcement learning algorithms

are susceptible to a lack of interpretability [158], making it difficult to understand or verify the learned decision-making process. Moreover, the performance of these algorithms is highly sensitive to the design of the reward function [159]. An unsuitable reward function may hinder the agent from learning an optimal policy.

2.4.3 Utility-Based Method

Utility-based methods aim to guide an agent's decision-making process by selecting actions that maximize its expected utility, which is typically formulated through reward or cost functions. A representative class of utility-based approaches is game theory.

Game theory, originally developed by Von Neumann and Morgenstern [160], provides a powerful mathematical framework for modeling and analyzing decision-making in scenarios involving conflicting interests among multiple individuals or agents. It centers on the study of strategic interactions, where each rational player aims to select the action that maximizes their own utility while considering the potential decisions of others [161]. Conventional game theory relies on the assumption of complete information, meaning that all participants are fully informed about the structure of the game [162]. This includes knowledge of the number of players involved, the set of available strategies for each player, and the payoff or utility outcomes associated with every possible combination of actions. Moreover, it presumes that all players are perfectly rational. They are capable of performing optimal reasoning based on this information and will always act to maximize their own utility. In this model, participants iteratively adjust their choices and beliefs until they reach a Nash equilibrium, which is the fundamental principle of traditional game theory [163]. A Nash equilibrium represents a steady state of a game where each player is reluctant to unilaterally change their strategy because such a change would not result in a more favorable outcome.

While game theory was initially applied in the field of economics, it is now being gradually introduced to the study of vehicle-vehicle or vehicle-pedestrian interactions in the context of AVs. A Zebra Crossing Game was introduced to explore the interaction between cyclists and vehicles in Norway, and the findings demonstrated that the real crossing behavior was consistent with the solution derived from game theory [164]. A game theory model, called the 'sequential chicken' model was proposed to simulate space competition between a vehicle and a pedestrian at an unsigned intersection [165]. This model also was extended in [166] by employing empirical data and the Gaussian Process to fit the model's parameters.

A static game model with complete information was proposed to resolve conflicts between AVs at the road intersection, aiming to help the vehicles pass through this area efficiently. This static model assumed that all involved vehicles are making decisions simultaneously [167]. On the other hand, a recent study developed a Stackelberg game model due to the belief that players usually make sequential decisions in the form of actions and responses rather than simultaneously during road conflicts [168]. This is a dynamic game model under the assumption that the vehicle and the pedestrian act as leaders, respectively. The leader initiates the decision, after which the follower selects an action to maximize its utility concerning the leader's strategy. Similar applications of the Stackelberg game model for simulating the interaction between vehicles and pedestrians are observed in references [169] and [14]. Figure 2.6 illustrates a representative example of this dynamic game framework, showing the sequential decision-making process between the vehicle (leader) and the pedestrian (follower).

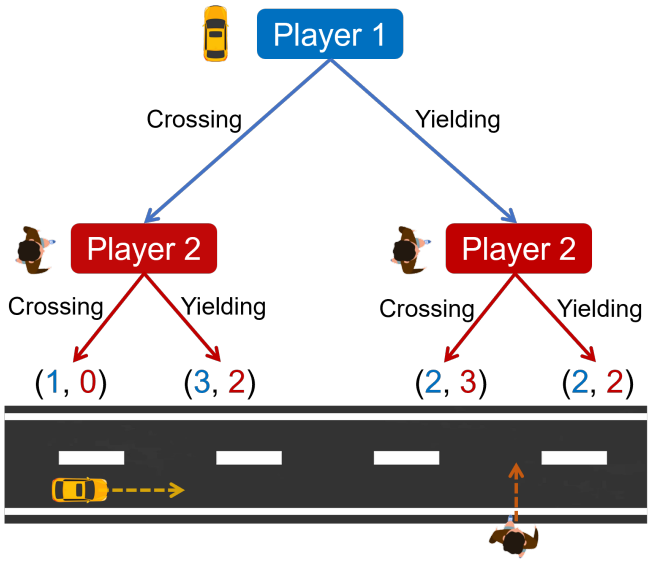


Figure 2.6: An illustrative example of a Stackelberg game with payoffs in an AV–pedestrian interaction scenario. The AV, represented as Player 1 in blue, acts as the leader and first decides whether to yield or cross. The pedestrian, shown as Player 2 in red, follows by observing the AV's action and subsequently selecting whether to cross or wait. The terminal nodes display the respective utilities for both agents, with blue indicating the AV's outcome and red reflecting the pedestrian's payoff.

The approaches mentioned earlier are all based on the conventional game theory model with players' complete rationality assumption. However, human behavior does not always conform to the predictions of the Nash equilibrium [170] in real-world situations due to bounded rationality and cognitive limitation [171]. To relax the assumption of complete rationality, Chen et al. [172] combined evolutionary game theory with cumulative prospect theory to formulate an interactive decision model at uncontrolled mid-block crosswalks. This method can simulate different behaviors within a pedestrian group but requires numerous parameters to complete the model fitting.

In contrast to classical game theory, behavioral game theory stands as a more accurate predictor of human behavior in real-world scenarios [173]. The findings of a comparison with four game theoretical models revealed that the behavioral game theoretical model outperforms the conventional model in forecasting interaction outcomes [174]. The researchers proposed that Nash equilibrium, normally with complete information may not sufficiently reflect the unpredictable actions of pedestrians at crosswalks. To simulate the joint behavior of pedestrians and vehicles, they provided a game theoretical framework, namely logit quantal response equilibrium [175], [176] with incomplete information, replacing Nash equilibrium. Moreover, level-k reasoning softened the perfect rationality assumption of Nash equilibrium by assuming that agents had different levels of reasoning [177]. The reasoning levels started at level 0, with k representing the depth of the agent's reasoning ability. An agent being at level k assumed that its opponent performed at level-(k-1) and made predictions based on this assumption. Level-k reasoning is frequently implemented in vehicle-vehicle interaction models in diverse scenarios, such as roundabout situations [178], lane changes [179], and intersections [180]. However, if the opponent's cognitive hierarchy is not at level-(k-1), the level-k model will not perform well in predicting its behavior. The study in [181] considered another game theoretical framework, known as the 'cognitive hierarchy' framework. This framework holds similarities to level-k but possesses special features for modeling vehicle-vehicle interaction. It allows the AV to interact with opponents of varying cognitive hierarchies, rather than being restricted to those only one level below it.

In conclusion, game theory can be a valuable framework for helping investigate the interactions between AVs and pedestrians. Compared to the rule-based approach, it is adept at handling more complicated and dynamic scenarios. Moreover, game theoretical models present better interpretability than learning-based approaches, as they explicitly model the strategic reasoning process of each agent. However, traditional game-theoretic frameworks often rely on strong assumptions of perfect rationality and complete information, which may not fully capture human behavior in real-world environments. Recent developments such as level-k reasoning and cognitive hierarchy models have been proposed to better approximate bounded rationality and improve the realism of decision-making models.

2.5 Human-Robot Interaction

Due to the limited research on AV navigation in dense pedestrian environments, many insights have been drawn from crowd navigation studies on mobile robots, whose interaction modeling techniques can offer useful inspiration for AV systems. Prior work in this domain falls into two categories: prediction-based planning or learning-based approaches.

Prediction-based planning methods include decoupled and coupled approaches. Decoupled methods first predict the surrounding agents' trajectories and then plan a collision-free path in the remaining free space [182]. However, they often lead to conservative behavior or freezing, as the predicted area may become overly occupied with no feasible path [183]. Coupled methods jointly plan trajectories for all agents by considering mutual interactions [184], but they require frequent updates when agents deviate from their planned paths, resulting in high computational cost and limited real-time applications [185]. Moreover, all prediction-based methods depend heavily on accurate pedestrian trajectory forecasts, which are often unreliable due to the unpredictability of human behavior.

Learning-based methods, particularly those based on deep reinforcement learning (DRL), have emerged as promising alternatives due to their ability to capture complex social interactions and support efficient online inference. To improve interaction modeling, many DRL frameworks for robot navigation have integrated advanced network architectures such as RNN [186], [187], [188], GNN [189], [190], [191], and attention mechanisms [192], [193], [194]. For example, [188] incorporated Long Short-Term Memory into DRL to encode observations from a variable number of surrounding agents into a fixed-length vector. A heterogeneous relational DRL framework was introduced to explicitly model diverse interaction types using a customized GNN [190]. In [193], a spatial graph was combined with an attention module to highlight the relative importance of nearby agents. While these approaches enhance interaction modeling, they often neglect the subjective uncertainty inherent in human cognition, such as prediction errors or biases that may arise when an agent attempts to infer pedestrian intentions or behaviors.

2.6 Chapter Conclusion

Through the discussion of relevant work in this chapter, it becomes evident that while autonomous driving has been extensively studied, most research on AV decision-making predominantly focuses on interactions with human-driven vehicles. In contrast, AV-pedestrian interactions remain relatively underexplored. This research gap persists

across both simplified scenarios involving a single pedestrian and more complex shared spaces where AVs need to navigate through crowds. Moreover, in existing AV–pedestrian interaction frameworks, pedestrian behavior is typically modeled in a pre-defined or reactive manner. Such approaches fail to capture the inherently interactive nature of real-world scenarios, where pedestrians continuously perceive, interpret, and respond to the AV's behavior in a bidirectional loop. This lack of adaptability and flexibility limits the realism and effectiveness of interaction models, especially in dynamic, uncertain environments. These limitations underscore the need for pedestrian models and AV decision-making frameworks that enable realistic interaction with pedestrians in dynamic and complex urban environments. These insights form the basis for the investigations presented in the following chapters.

Chapter 3

Virtual Reality Experiment Study for Vehicle-Pedestrian Interaction

3.1 Background

This chapter explores the critical factors influencing the decision-making processes of human drivers and pedestrians respectively by using virtual reality technology. The goal is to gain deeper insights into how these road users make interactive decisions in traffic scenarios, particularly in situations involving potential conflicts such as pedestrian crossings.

A review of the literature reveals numerous studies that investigate interactive behaviors between human drivers and pedestrians and create models for predicting pedestrian crossing intentions [195], [196]. However, most studies have primarily focused on identifying variables that explain pedestrian crossing behavior rather than predicting pedestrian crossing intention. These parameters may not be obtained prior to making a decision. For example, frequently used indicators such as waiting time are usually measured after a decision has been made [39]. Moreover, decision-making models in these studies often rely on certain factors that may not be readily available in real-world situations. These factors such as age [32], sensation seeking [33], and social value orientation [34] are typically gathered through post-experiment questionnaires or interviews and are often inaccessible or impractical for real-time use in autonomous vehicle systems.

To address these limitations, interactive vehicle-pedestrian experiments were conducted using virtual reality technology, and behavior analysis was performed with an exclusive focus on the kinematic perspective, since the kinematic variables are easily accessible prior to AVs' decision-making in real-world scenarios. Specifically, virtual reality trials were car-

ried out to collect data based on the state-of-the-art distributed driving-pedestrian simulator platform, investigating the most important factors impacting human drivers' and pedestrians' decision-making processes, and analyzing the inherent relationships between motion parameters and pedestrians' behavior as well as driver-yielding behavior.

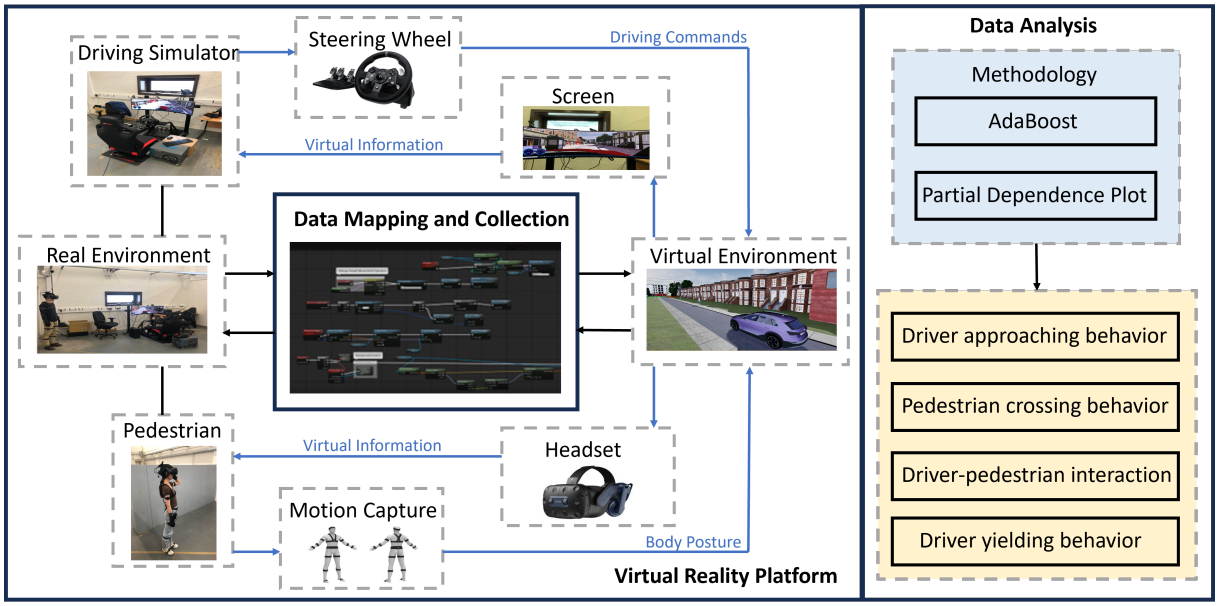


Figure 3.1: The theoretical framework of the virtual reality study.

Fig. 3.1 shows the theoretical framework of our study. The virtual reality platform can create an immersive virtual environment where vehicles and pedestrians can interact. A driving simulator, equipped with a steering wheel and a screen, is used to emulate the driving experience for a human participant acting as the vehicle driver. The screen renders the visual feedback of the virtual environment, while the steering wheel captures driving commands, such as acceleration, braking, and steering inputs. These commands are then transmitted to the virtual environment to control the behavior of the simulated vehicle. Simultaneously, a real human pedestrian participant is situated in the same environment, equipped with a VR headset and a motion capture system. The headset provides an immersive view of the shared virtual space, while the motion capture system records body posture and walking behavior. This information is used to animate the movement of the virtual pedestrian within the simulation, ensuring that their actions accurately reflect natural human motion. During experiments, the data collection module records all relevant data for subsequent analysis. On the right side of the framework, the data analysis pipeline is depicted. This phase involves the application of machine learning techniques, specifically the AdaBoost algorithm, to interpret behavioral patterns extracted from the collected data. To enhance interpretability, Partial Dependence Plots are employed to visualize the influence of individual features on model predictions. The combined analytical approach supports detailed investigation of AV-pedestrian interaction.

3.2 Methodology

3.2.1 AdaBoost

AdaBoost, short for Adaptive Boosting, stands as a potent ensemble learning technique employed in classification tasks. The core principle of this approach involves aggregating weak learners to form a robust learner, a fundamental aspect of the broader 'Boosting' methodology within ensemble learning [197]. Despite these weak learners only slightly outperforming random guessing in terms of prediction accuracy, their learning potential is pooled to create a more trustworthy model. Unlike bagging methods that train base learners in parallel [198], AdaBoost employs a sequential strategy in which the performance of each learner informs the training of the next, specifically by reweighting the training samples to emphasize those that were previously misclassified.

In the AdaBoost algorithm, the process commences by creating a weak classifier, where the weights assigned to the training samples are identical. Formally, for a training set of N samples, the initial weight distribution is defined as:

$$D_1(i) = \frac{1}{N}, \ \forall i \in \{1, \ldots, N\} \tag{3.1}$$

where $D_1(i)$ denotes the initial weight assigned to the i-th training sample. Subsequently, these sample weights are adjusted based on the performance of the proceeding weak classifier, a crucial step for the successive classifier and its progress [199]. This leads to an amplification in the weight of samples that were erroneously classified and a reduction in correctly classified samples. The classification error of the current weak learner is computed as the weighted sum of the misclassified instances [199]:

$$\varepsilon_t = \sum_{i=1}^{N} D_t(i) \cdot \mathbb{I}[h_t(x_i) \neq y_i] \tag{3.2}$$

where ε_t is the weight error of the weak classifier h_t at iteration t, $D_t(i)$ is the current weight of sample i, x_i is the input, $y_i \in \{-1, 1\}$ is the true label, and $\mathbb{I}[\cdot]$ is the indicator function that returns 1 if the prediction is incorrect, and 0 otherwise. The importance of each weak classifier is quantified by a weight assigned to it based on its accuracy, defined as:

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \tag{3.3}$$

where α_t indicates the contribution of the t-th weak classifier to the final ensemble. A smaller ε_t results in a larger α_t, giving the corresponding weak classifier more influence in the final prediction. Using the updated sample weights, another weak classifier is constructed. The new weights are computed according to:

$$D_{t+1}(i) = \frac{D_t(i) \cdot e^{-\alpha_t y_i h_t(x_i)}}{Z_t} \tag{3.4}$$

where Z_t is a normalization factor to ensure that the updated weights form a valid probability distribution:

$$\sum_{i=1}^{N} D_{t+1}(i) = 1 \tag{3.5}$$

If a sample x_i is correctly classified, the exponential factor in the weight update becomes $e^{-\alpha_t}$, resulting in a decrease in ints weight. Conversely, if the sample is misclassified, leading to a higher weight. As the process iterates, subsequent weak classifiers prioritize samples that were previously misclassified, progressively refining the model's performance. Finally, a strong learner is produced by combining these sequentially weak classifiers. The overall prediction is given by a weighted majority vote:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{3.6}$$

where $H(x)$ is the final output of the AdaBoost classifier and the sign function determines the final class label by aggregating the weighted votes of all weak learners. The determination of each weak classifier's contribution to the strong learner hinges on its classification accuracy, assigning greater weights to more accurate classifiers. AdaBoost algorithm efficiently utilizes the collective insights from these weak classifiers, leveraging their combined strength and optimizing the model's predictive capabilities through the cascading impact of these weak classifiers [200]. The detailed procedure of the AdaBoost algorithm is presented in Algorithm 1.

AdaBoost has demonstrated exceptional effectiveness and versatility in various applications, whose adaptability, capacity to handle intricate datasets, and ability to improve predictive accuracy make it a popular choice in the machine-learning community. In this study, we adopt the AdaBoost algorithm to analyze behavioral data collected from virtual reality experiments involving human drivers and pedestrians. AdaBoost is particularly well suited to this analysis for several reasons. First, it provides explicit and interpretable feature-importance measures, enabling a transparent understanding of the relative contributions of different factors to interaction outcomes. Second, AdaBoost performs reliably on relatively small datasets, such as our VR-based dataset. Finally, its ensemble structure combing multiple weak learners enables AdaBoost to achieve stable and robust perform-

Algorithm 1: AdaBoost Algorithm

Input: Training data $\{(x_i, y_i)\}_{i=1}^{N}$, number of iterations T
Output: Strong classifier $H(x)$
Initialize sample weights $D_1(i) = \frac{1}{N},\ \forall i \in \{1, \ldots, N\}$;
for *iteration* $t = 1$ **to** T **do**
    Train weak classifier $h_t(x)$ using weighted distribution D_t;
    Compute error $\varepsilon_t = \sum_{i=1}^{N} D_t(i) \cdot \mathbb{I}[h_t(x_i) \neq y_i]$;
    Compute classifier weight $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$;
    Update sample weights:
        $D_{t+1}(i) = \frac{D_t(i) \cdot e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$, where Z_t is a normalization factor;
Combine weak learners into final strong classifier:
    $H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$;
return $H(x)$;

ance in the presence of variability and noise in human-centered data. These characteristics make it an appropriate and effective choice for our task. By applying AdaBoost to classify behavioral outcomes and rank feature importance, we aim to identify the key factors influencing decision-making under interaction scenarios.

3.2.2 Partial Dependence Plot

The Partial Dependence Plots (PDPs), introduced by Friedman [201], effectively illustrate the marginal impact of one or two features on the predicted outcomes of a machine learning model. These plots showcase the functional relationship between a feature variable and the model's prediction results. This relationship is determined by applying a set of data to the machine learning model, altering the value of the feature of interest while keeping supplementary features constant, and then analyzing the resulting model output. This analysis helps ascertain whether the relationship is roughly linear, monotonic, or more complex [202]. PDPs offer valuable advantages, as they provide clear and understandable insights into the features' influences. To further elaborate, a partial dependence function is defined as follows [203]:

$$f_S(X_S) = E_{X_S} f(X_S, X_C) \tag{3.7}$$

where X_S is a subset of the target feature, and X_C is a subset of all features excluding the feature of interest. $f(\cdot)$ denotes the predictive function of the machine learning model, and $\mathbb{E}_{X_C}$ represents the marginal expectation over the distribution of the remaining featrues X_C. The function $f_S(X_S)$ thus captures the average predicted response as a function of X_S, holding all other featrues marginalized.

Normally, X_S consists of one or two attributes. The PDP takes a one-dimensional form when X_S contains a single feature, effectively illustrating the impact of that specific feature on the expected output. However, when X_S contains two features, the PDP transforms into a two-dimensional representation. This version not only showcases the influence of a single feature on the predicted output but also reveals the interactive relationship between the two features and its impact on the predicted results [204]. All training data can be averaged to estimate the partial dependence function as [203]:

$$\bar{f}_S(X_S) = \frac{1}{N}\sum_{i=1}^{N} f(X_S, x_{iC}) \tag{3.8}$$

where N is the number of samples in the training dataset, and x_{iC} denotes the true observed values of the complementary feature set X_C for the i_{th} training instance.

3.3 Experiment

3.3.1 Participants

This study recruited a total of 32 participants, including 16 drivers and 16 pedestrians, from various genders and age groups to participate in our simulation experiment. The choice of 32 participants was made with considerations of feasibility and the anticipated effects in our experimental design. All participants were either students or teachers, who were randomly divided into 16 groups, each consisting of both a driver and a pedestrian. To ensure the effectiveness of the experiment, participants were required to meet certain criteria. They were in good health with no vision or motion impairments. Additionally, all drivers possessed a valid UK driving license with a minimum of 3 years of driving experience. Prior to their involvement in the experiment, participants were given a comprehensive briefing on the experimental procedures, familiarized with the experimental equipment, and requested to provide informed consent.

3.3.2 Appartus

Due to the costly and time-consuming data collection in real-world physical experiments, virtual reality technology was applied in our investigation to gather data on vehicle-pedestrian interactions.

(a) (b)

Figure 3.2: A driver-pedestrian simulator platform. (a) Driving simulator integrated external devices. (b) Pedestrian simulator.

Our driving-pedestrian simulator was developed using Unreal Engine (UE), a highly advanced and versatile 3D creative tool compatible with various external devices, including virtual reality headsets, steering wheels, and motion capture systems. Initially, we designed a synthetic urban environment using Unreal Engine 4.26 to facilitate real-time interactions between drivers and pedestrians. The simulator integrated several key components, including the Logitech G920 steering wheel, VIVE Pro 2 headset, and Perception Neuron Motion Capture System, alongside other sensors and external devices (see Fig. 3.2). Upon full setup of the simulation system, the driver seated in front of the driving simulator was presented with a realistic view of the interior of the vehicle on a 49-inch curved screen (see Fig. 3.3). The Logitech G920 steering wheel effectively captured the driver's inputs for steering, throttle, and braking, which was then transmitted to control the virtual vehicle developed using Python. The Python script captures driving commands and updates the vehicle's kinematic information within the Unreal Engine 4.26.

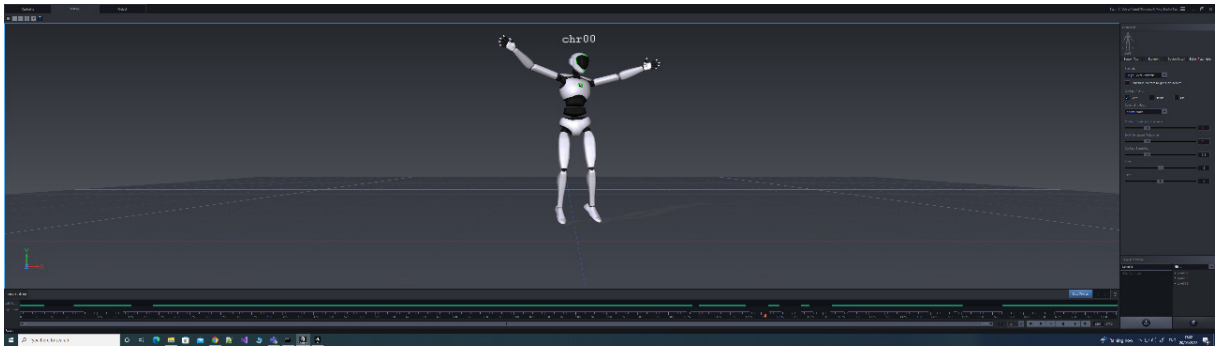
Meanwhile, the motion capture system and VR headset afforded pedestrians a virtual perspective within the simulated environment, enabling the capture of posture data. The motion capture system transmits posture information to the UE, and concurrently, the UE sends the view of the virtual pedestrian back to the VR headset. For the Perception Neuron Motion Capture System, a software named Axis Neuron is employed

Figure 3.3: Realistic view of the interior of the vehicle on a 49-inch curved screen

to handle data from motion capture sensors, which can be configured through USB in single-performer mode or Ethernet in multi-performer mode. In our project, we utilize Ethernet for data transformation. Axis Neuron also enables real-time capture, recording, and streaming into Unreal Engine. The appearance of the pedestrian in the virtual reality environment of the real-time capture in Axis Neuron and data live streaming in UE are depicted in Fig. 3.4. Indeed, both motion capture sensors and VR headsets can offer precise position and orientation data for the pedestrian's head. Yet, the VR headset can introduce interference with the motion capture unit worn on the head. As a result, we prioritize relying on the position and orientation data provided by the VR headset to accurately map the pedestrian within the virtual environment. The motion capture system is specifically employed for animating the avatar's body posture. This data was subsequently translated into virtual pedestrian movements, allowing for the seamless integration of real-time interactions between drivers and pedestrians.

3.3.3 Data Collection

To capture real-time driver-pedestrian interactions, we designed a single-lane road spanning 3.65 meters wide within a virtual environment. It's important to emphasize that all the elements in the virtual environment, such as pedestrians, vehicles, buildings, and trees, are precisely adjusted to a 1:1 scale with the real world. This adjustment holds significant importance as it profoundly influences the perception of speed for both the



(a)

(b)

Figure 3.4: The appearance of the pedestrian in the virtual reality environment. (a) Real-time capture in Axis Neuron. (b) Data live streaming in the Unreal Engine.

driver and pedestrians, as depicted in Fig. 3.5. The benefit of designing a loop-shaped road is that it allows the driver to maintain continuous driving without the need to restart the simulation at each episode. In this scenario, drivers navigated continuously along a circular road at random speeds, reacting to pedestrians appearing in their field of vision by either yielding or continuing their course. Pedestrians were randomly spawned on the opposite side of the circular road and decided whether to cross the road when a car approached. In cases where pedestrians opt not to yield, they proceed to cross the road directly. The pedestrian was also unaware of when the car would appear. Fig. 3.6 and Fig. 3.7 illustrate one of the interactive scenarios in the real world and virtual reality, respectively. The fact that unpredictability of when pedestrians would appear introduced a significant level of uncertainty, fostering more genuine and lifelike driver responses. Our study focused on the interactions between a single pedestrian and a single driver, with a total of 32 participants, encompassing diverse ages and genders, grouped into 16 pairs. These experiments were intentionally designed as a scenario between a single vehicle and a single pedestrian to allow a controlled investigation of the key behavioral factors influencing pedestrian and human driver decision-making. A more complex multi-agent setup would introduce multiple interacting variables and confounding effects, making it difficult to isolate the fundamental mechanisms of vehicle-pedestrian interaction. Each pair will conduct the experiment for an hour, with a 20-minute break every half-hour. Consequently, we recorded one hour of data for each group. However, as some instances

lacked meaningful interaction between pedestrians and vehicles, our study developed a program to detect and filter out these invalid data based on factors such as the positions of pedestrians and vehicles, as well as their relative distances. Each pair, on average, contributed 30 minutes of driver-pedestrian interaction effective data recorded after the removal of invalid data. In total, we amassed a total of 8 hours' worth of driver-pedestrian trajectory data for analysis. Specifically, we gathered kinematic data pertaining to driver-pedestrian interactions, mainly including the absolute position of both the vehicle and the pedestrian, the vehicle's speed, and the driver's inputs related to steering, throttle, and braking inputs at each time step. It is noted that the motion capture system and VR headset have their own reference coordinate system respectively. Consequently, it is necessary to map their data into the world coordinate in the UE.

3.4 Results and Discussion

In this section, we conducted an analysis of drivers' approaching behavior, pedestrians' crossing behavior, as well as their interactions and drivers' yielding behavior within specific ranges using collected data.

3.4.1 Drivers' Approaching Behavior

This part delves into the critical factors influencing drivers' decision-making process when encountering pedestrians. Table 4.2 provides an overview of specific variables considered, primarily focusing on vehicle dynamics. In this analysis, the crossing point is defined as the juncture where the vehicle and pedestrian trajectories intersect [205], as shown in Fig. 3.8. The initial time is defined as the moment a pedestrian enters a driver's field of vision. Previous research has demonstrated that a driver's decision time initiates upon first visual contact with the pedestrian and ends when they either engage the pedal or manipulate the steering wheel [206]. Consequently, this study posits that when a driver yields to a pedestrian, the driver's decision time corresponds to when to hit the brake or release the accelerator pedal. Conversely, if the driver crosses the crossing point before the pedestrian, the decisive moment centers on when to depress the accelerator pedal or release the brake pedal.

AdaBoost, which provides accurate results without the risk of overfitting [197], is used to analyze the factors influencing a driver's decision-making process. The output of the AdaBoost model is whether the driver yields to the pedestrian or not. To quantify the driver's decision-making, we categorize the scenario where the driver crosses the crossing

(a)

(b)

Figure 3.5: The virtual urban environment. (a) Overview of the virtual environment. (b) 3.65m single lane road.

(a)

(b)

Figure 3.6: Data collection with driving-pedestrian simulators in the real world

(c)

Figure 3.6: (Continued) Data collection with driving-pedestrian simulators in the real world

Figure 3.7: One of the vehicle-pedestrian interactive scenarios in virtual reality

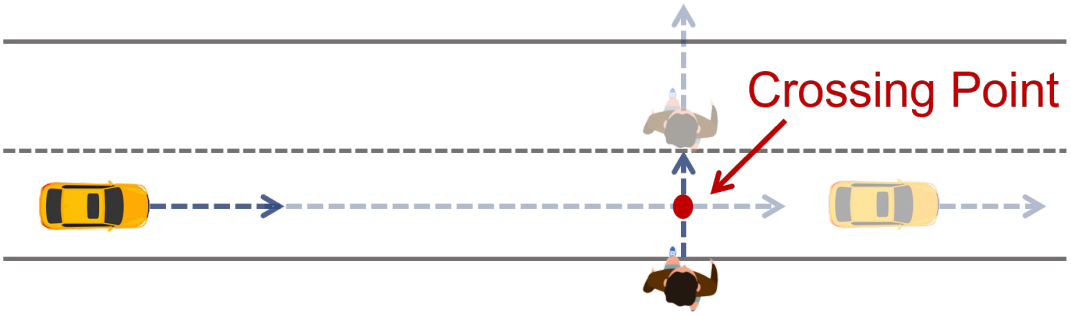


Figure 3.8: Illustration of vehicle-pedestrian crossing point.

Variable	Description	Unit
init_TTA	The initial time to arrival, representing the time for the vehicle to reach the crossing point when a pedestrian enters a driver's field of vision	s
init_v_veh	The vehicle speed when a pedestrian enters a driver's field of vision	m/s
init_d_longi	The distance for the vehicle to the crossing point when a pedestrian enters a driver's field of vision	m
init_d_lat	The distance for the pedestrian to the crossing point when a pedestrian enters a driver's field of vision	m
TTA	The time to arrival, representing the time for the vehicle to reach the crossing point when a driver is at his decision time	s
v_veh	The vehicle's speed when a driver is at his decision time	m/s
a_veh	The vehicle's acceleration when a driver is at his decision time	m/s^2
d_longi	The distance for the vehicle to the crossing point when a driver is at his decision time	m
d_lat	The distance for the pedestrian to the crossing point when a driver is at his decision time	m
v_ped	The pedestrian's speed when a driver is at his decision time	m/s

Table 3.1: Variables used for analyzing drivers' approaching behavior.

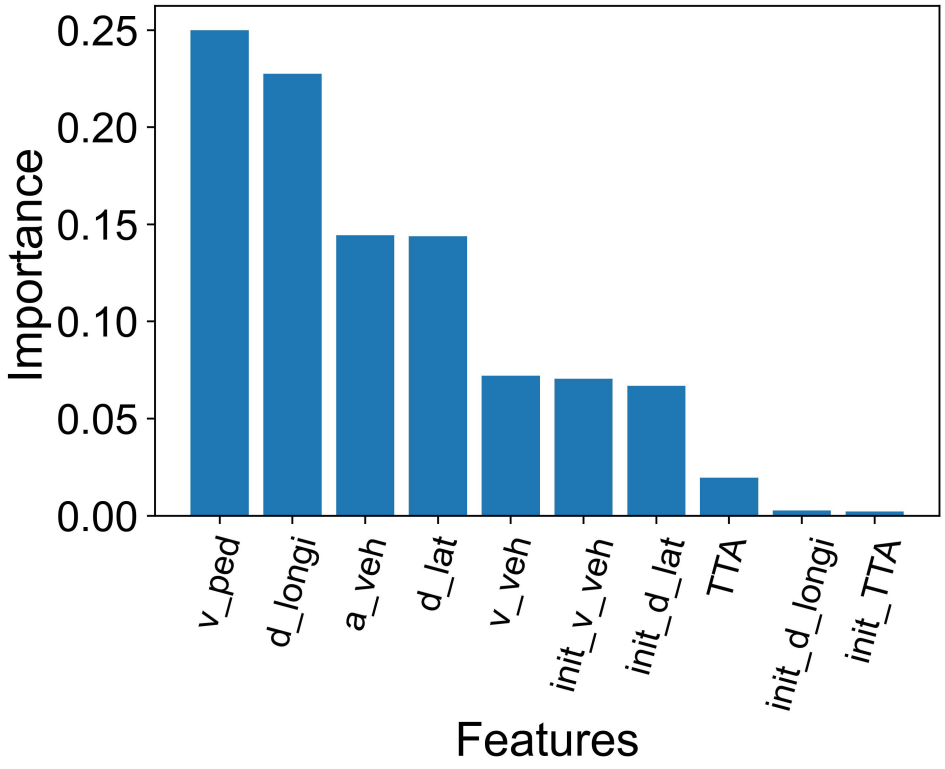


Figure 3.9: Feature importance in driver's decision-making process.

point first as the decision not to yield. Conversely, if the pedestrian crosses the road first, we interpret it as the driver choosing to yield to the pedestrian. The model was trained on a total of 479 samples. Among these, 241 samples correspond to scenarios where the driver yielded, while 238 samples represent instances where the driver proceeded without yielding. To develop the model, 80% of the dataset is randomly selected for training purposes, while the remaining 20% is reserved for evaluating the model's performance. The optimal parameters of the model (*max-depth=2, n-estimators=12, learning-rate=0.5*) are determined through a five-fold cross-validation process. The model's accuracy on the test set is 0.927, accompanied by an *f1-score* of 0.932.

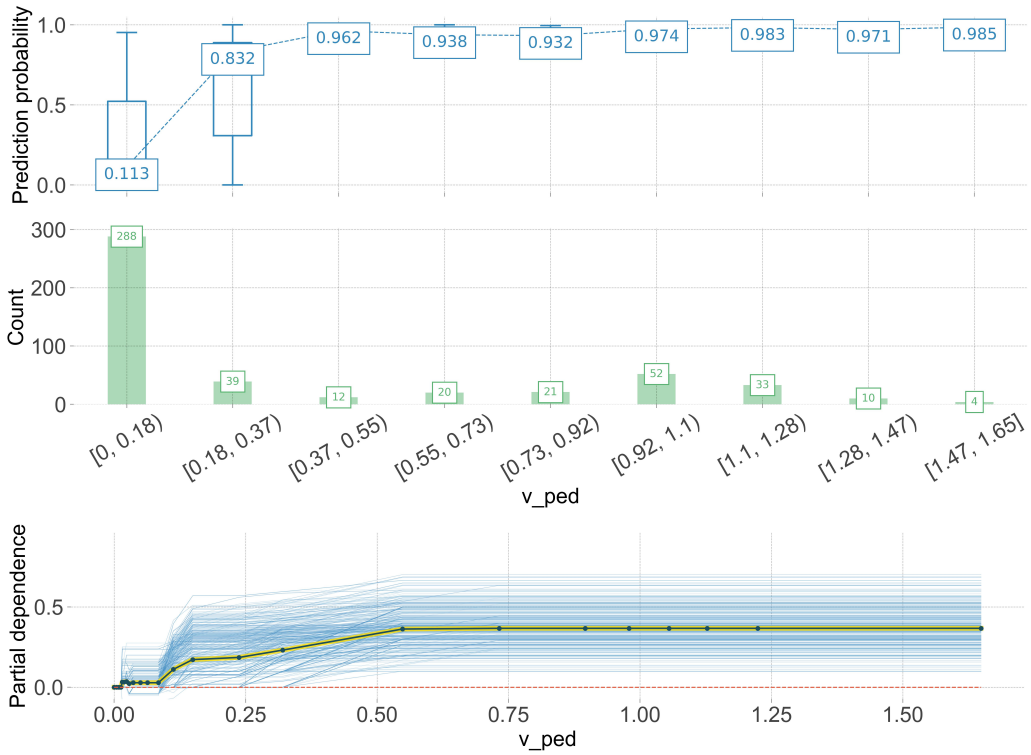
In Fig. 3.9, the importance of variables in drivers' decision-making is displayed. The more significant a variable is, the greater its impact on the driver's decision-making process. It can be seen from Fig. 3.9 that the most crucial factor is the pedestrian's speed, followed by the distance between the vehicle and the crossing point at the moment the driver makes their decision. Moreover, it is evident from observations that specific factors exert a substantial yet non-linear impact on a driver's decision-making process. It is also worth noting that variables at the decision moment hold more significance than those at the initial moment. Thus, the analysis will mainly focus on the relevant parameters at the drivers' decision-making moment.

The PDP effectively illustrates how one or two variables affect the machine learning model's prediction results. It can reveal a linear, monotonic, or more intricate relationship between variables and predictions [207]. PDPs make it easy to observe the impact of the independent variable on the drivers' decision-making process. The PDPs for relative variables in drivers' decision time are presented in Fig. 3.10, Fig. 3.11 and Fig. 3.12.

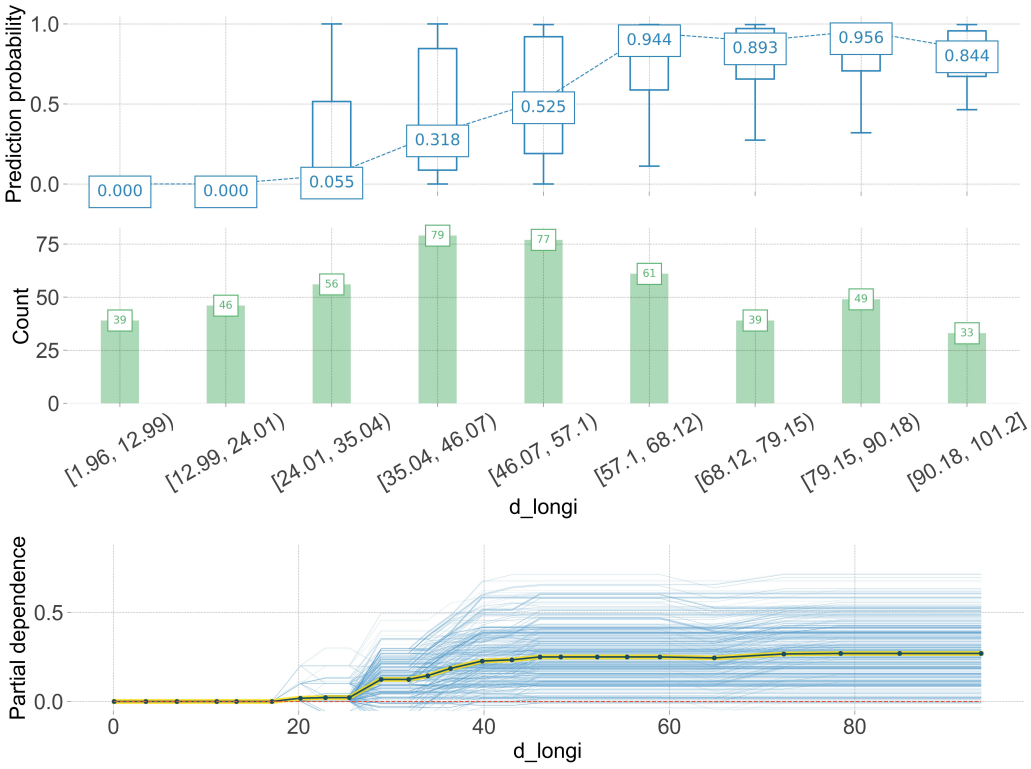
It is imperative to emphasize that increasing pedestrian walking speed from 0 to 0.5 m/s sharply elevates the probability of vehicles yielding to them. This occurs because changes in pedestrian speed indicate their intention to cross the road, prompting the driver to decide to yield. The partial dependence value also escalates during this speed range before stabilizing, signifying that 0.5 m/s is the threshold at which the driver's probability of yielding improves if the pedestrian walks more quickly, beyond this point, its effectiveness diminishes. Therefore, pedestrian speed plays an indispensable role in a driver's decision-making process.

As the TTA extends from 2 to 6 seconds, there is a notable increase in the likelihood of drivers yielding to pedestrians. When the TTA exceeds 6 seconds, the probability of drivers yielding is approximately 0.9. Similarly, the distance from the vehicle to the crossing point follows a comparable pattern with TTA. Moreover, TTA and the longitudinal distance have positive partial dependence values, indicating that their increasing effects on the driver's yielding are favorable. As the TTA and the longitudinal distance continue to increase, pedestrians' tendency to cross the road becomes more prominent, facilitating drivers in perceiving their intentions and consequently amplifying the probability of drivers yielding to pedestrians. These findings align with previous research on the topic [26, 37].

Although the likelihood of the driver yielding increases with higher vehicle speed, the partial dependence plot shows a weak relationship between speed and the driver's decision-making process. Furthermore, if a vehicle decelerates at a rate of exceeding 0.75 m/s^2, the driver is more likely to yield. However, if the vehicle is accelerating, the probability of the driver's yielding decreases significantly. Similarly, the driver's yielding probability drops as the pedestrian is farther away from the crossing point. This is because when the distance between the pedestrian and the intersection is greater, the pedestrian requires more time to cross the road, resulting in a higher probability of not crossing. Consequently, the driver is less likely to yield. Additionally, the partial dependence values for increased vehicle acceleration and lateral distance are negative, indicating that increases in these factors have a detrimental effect on the driver's ability to yield.

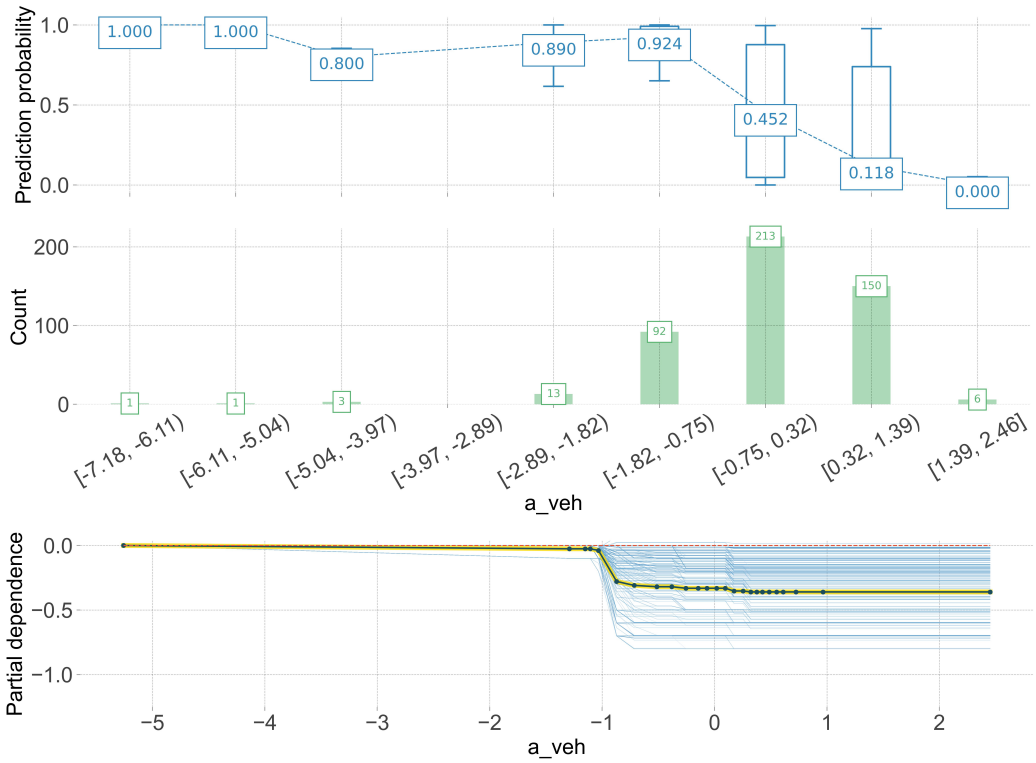


(a) Pedestrian's speed

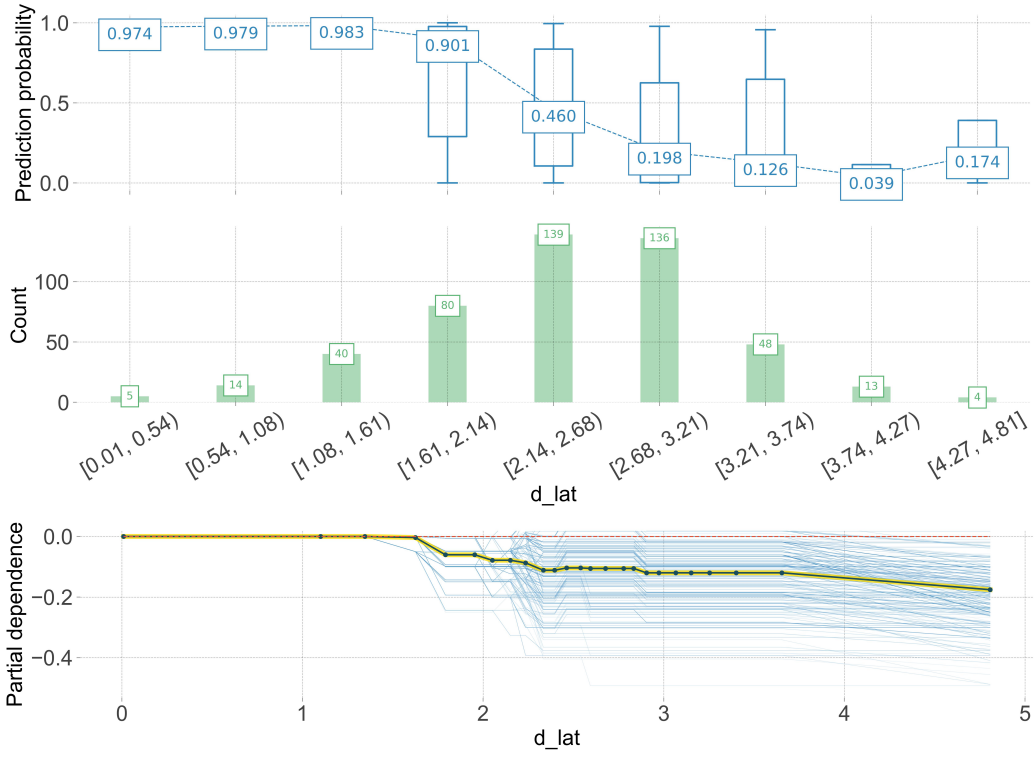


(b) Longitudinal distance

Figure 3.10: PDPs for the pedestrian's speed and the longitudinal distance in the driver's decision-making process.

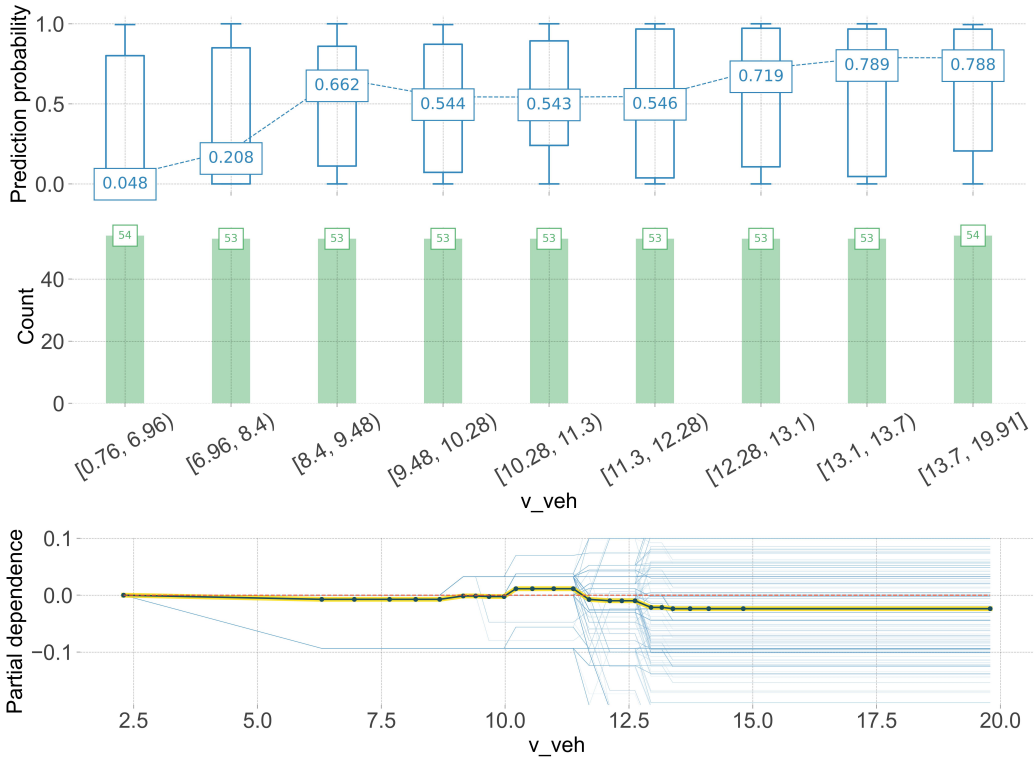


(a) Vehicle's acceleration

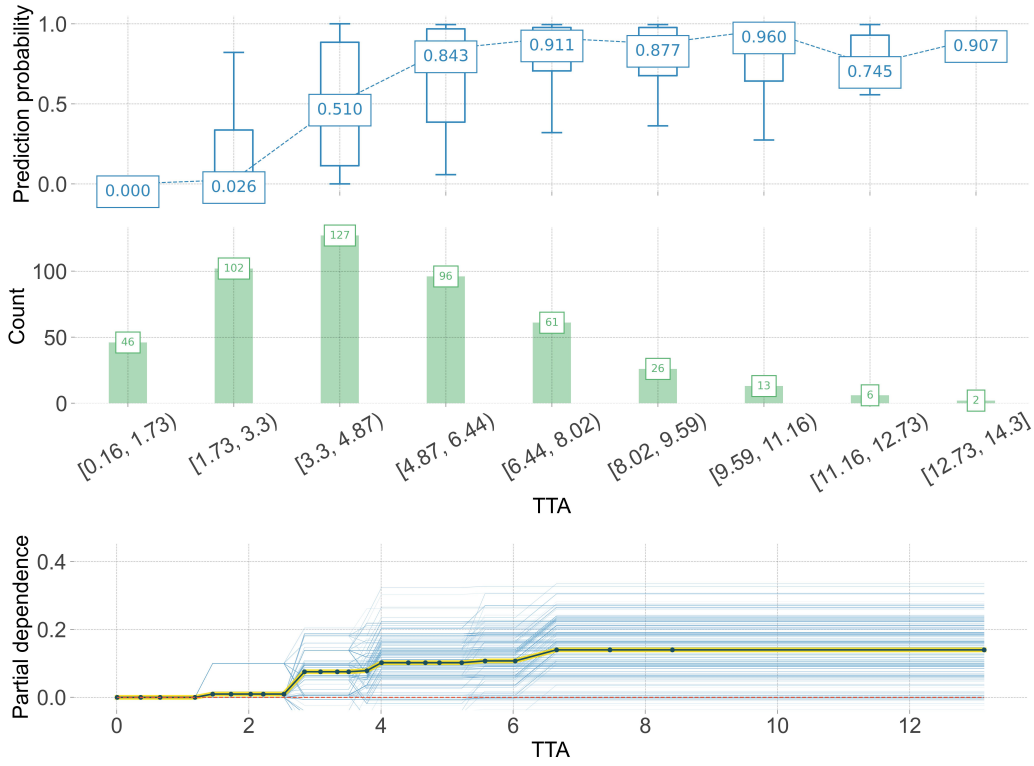


(b) Lateral distance

Figure 3.11: PDPs for the vehicle's acceleration and the lateral distance in the driver's decision-making process.



(a) Vehicle's speed



(b) TTA

Figure 3.12: PDPs for the vehicle's speed and the TTA in the driver's decision-making process.

3.4.2 Pedestrians' Crossing Behavior

This part explores the factors influencing a pedestrian's decision to cross the road in the presence of an approaching vehicle. The explanatory variables under consideration are outlined in Table 3.2. The initial moment is triggered when the pedestrian spots the approaching vehicle for the first time. Additionally, this study posits that if a pedestrian chooses to cross the road, their decision-making time is when their rate of speed change exceeds 1 m/s^2. Conversely, if the pedestrian decides to yield to the vehicle, their decision-making time is averaged based on the decision-making times for road crossing.

Our study also applies the AdaBoost approach to analyze the diverse factors that influence pedestrians when crossing the road. During the model development stage, we randomly select 80% of the dataset to train the model, reserving the remaining 20% of the data for model evaluation. Through a five-fold cross-validation process, we establish the optimal parameters of the model (*max-depth=4, n-estimators=61, learning-rate=0.3*). The model achieves an accuracy of 0.935 and an *f1-score* of 0.935 on the test set.

Fig. 3.13 illustrates the significance of variables in influencing pedestrian crossing decisions, with the impact increasing in proportion to the variable's importance. The figure highlights that the distance between the vehicle and the crossing point when a pedestrian is making a decision holds the utmost significance, followed by the vehicle's acceleration. On the contrary, the initial vehicle speed and initial TTA are relatively less critical. Notably, it is observed that the factors with the greatest impact exhibit a non-linear influence on the decision-making process of pedestrians. This study will also provide a detailed analysis through the partial dependence plots.

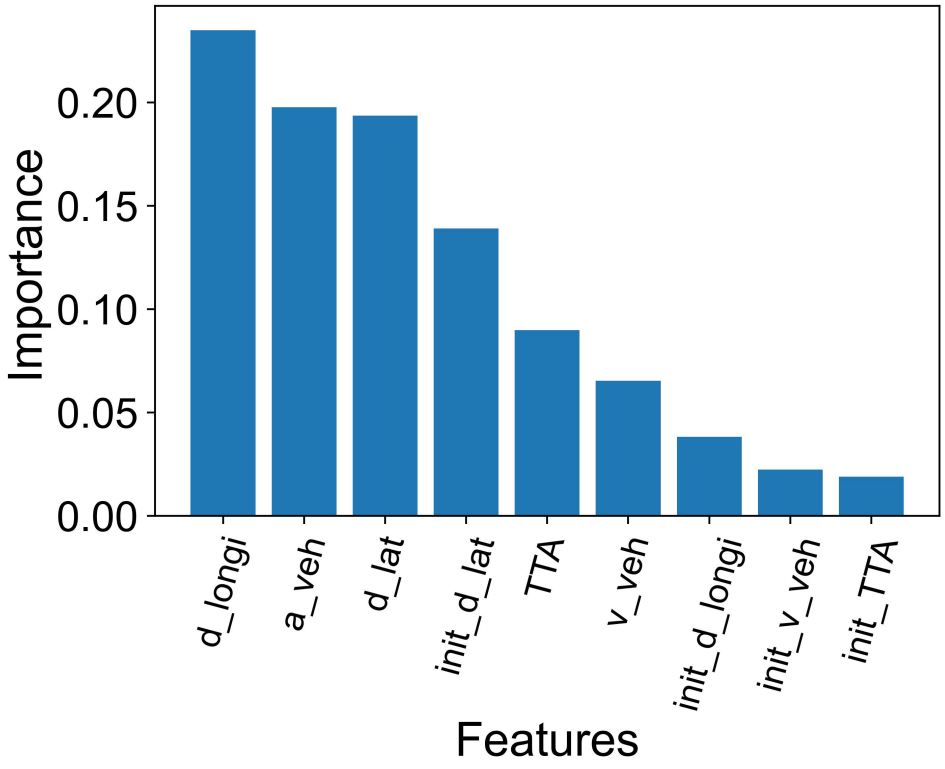


Figure 3.13: Feature importance in the pedestrian's decision-making process.

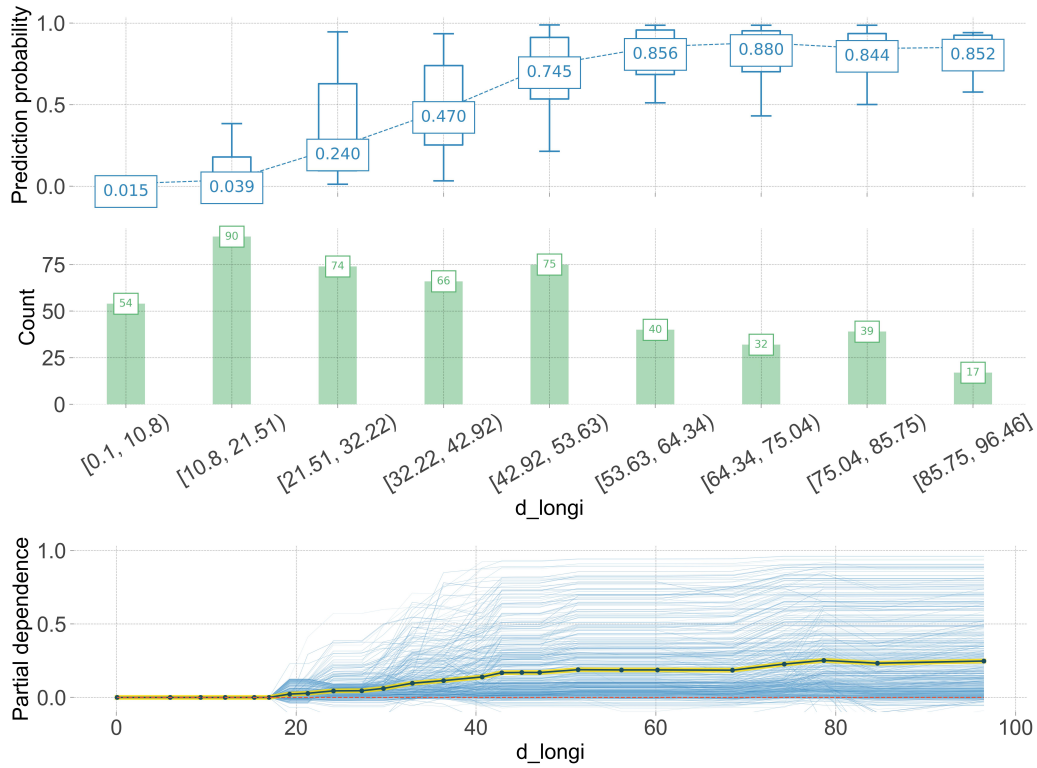
Variable	Description	Unit
init_TTA	The initial time to arrival, representing the time for the vehicle to reach the crossing point when a vehicle enters the pedestrian's field of vision	s
init_v_veh	The vehicle speed when a vehicle enters the pedestrian's field of vision	m/s
init_d_longi	The distance for the vehicle to the crossing point when a vehicle enters the pedestrian's field of vision	m
init_d_lat	The distance for the pedestrian to the crossing point when a vehicle enters the pedestrian's field of vision	m
TTA	The time to arrival, representing the time for the vehicle to reach the crossing point when the pedestrian is making a decision	s
v_veh	The vehicle's speed when the pedestrian is making a decision	m/s
a_veh	The vehicle's acceleration when the pedestrian is making a decision	m/s^2
d_longi	The distance for the vehicle to the crossing point when the pedestrian is making a decision	m
d_lat	The distance for the pedestrian to the crossing point when the pedestrian is making a decision	m

Table 3.2: Variables used for analyzing pedestrian crossing behavior.

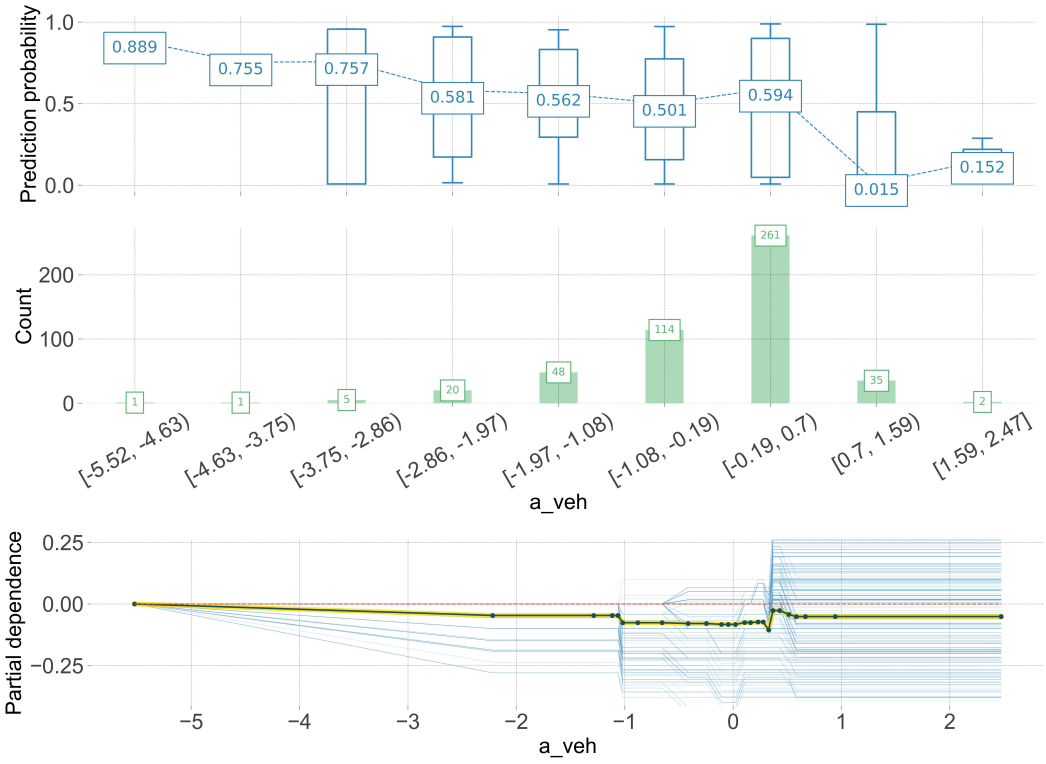
The information presented in Fig. 3.14, Fig. 3.15 and Fig. 3.16 demonstrates the impact of relevant variables on pedestrians' decision to cross the road. It is revealed that the longer distance for the vehicle to the crossing point exhibits an increasing crossing probability trend. Longer longitudinal distances encourage pedestrians to cross before vehicles approach. The partial dependence plot suggests that the probability of pedestrians crossing the road remains relatively steady when d_longi is under 20 meters. However, beyond this point, positive values that accompany the increase in distance strongly suggest a positive correlation between distance and propensity for pedestrian crossings. Conversely, as the lateral distance increases, the pedestrian crossing probability drops, as evidenced by the negative partial dependence values. This could possibly be attributed to the longer time required for pedestrians to cross the road with the larger lateral distance, thereby reducing the probability of crossing. Moreover, TTA is likely to have a similar effect as d_longi on pedestrian crossing decisions.

According to the partial dependence plot, it is evident that an increase in vehicle speed has an adverse effect on pedestrians who are crossing the road. However, the predicted probability shows fluctuations with rising vehicle speed. The probability initially decreases, but then it rises again. Notably, if the vehicle speed exceeds 12 m/s, the predicted probability of pedestrians crossing the street surpasses 0.5. Further investigation revealed that this phenomenon is linked to the experimental setup, where higher speeds typically coincide with longer longitudinal distances. Pedestrians are more inclined to cross the road with a larger longitudinal distance. It's important to emphasize that this PDP diagram solely represents the marginal effect of the vehicle speed variable on pedestrian crossing decisions and does not account for the interplay between vehicle speed and longitudinal distance. Another aspect unveiled by the partial dependence plot is that an increase in acceleration generally reduces the likelihood of pedestrians crossing the road. Pedestrians are more likely to cross when the vehicle acceleration is lower than -1 m/s^2.

Based on the analysis above, it becomes evident that pedestrians' decision to cross the road is influenced not only by a single variable but also by the interaction of two or more variables. Fig. 3.17 illustrates the interactive effects of vehicle speed and longitudinal distance, as well as vehicle acceleration and longitudinal distance. In Fig. 3.17a, a discernible correlation is observed between the likelihood of pedestrians crossing the road and both the longitudinal distance and vehicle speed. With a longitudinal distance lower than 35 meters, the vehicle's speed primarily affects pedestrian crossing behavior. However, for longer longitudinal distances, both the distance and the vehicle's speed play a role in shaping pedestrian behavior. Specifically, an increase in distance and speed amplifies the likelihood of pedestrian crossings, whereas a decrease in distance and speed diminishes this probability. Similarly, in Fig. 3.17b, a rise in both longitudinal distance and vehicle acceleration corresponds to an increased probability of pedestrian road crossings. Con-

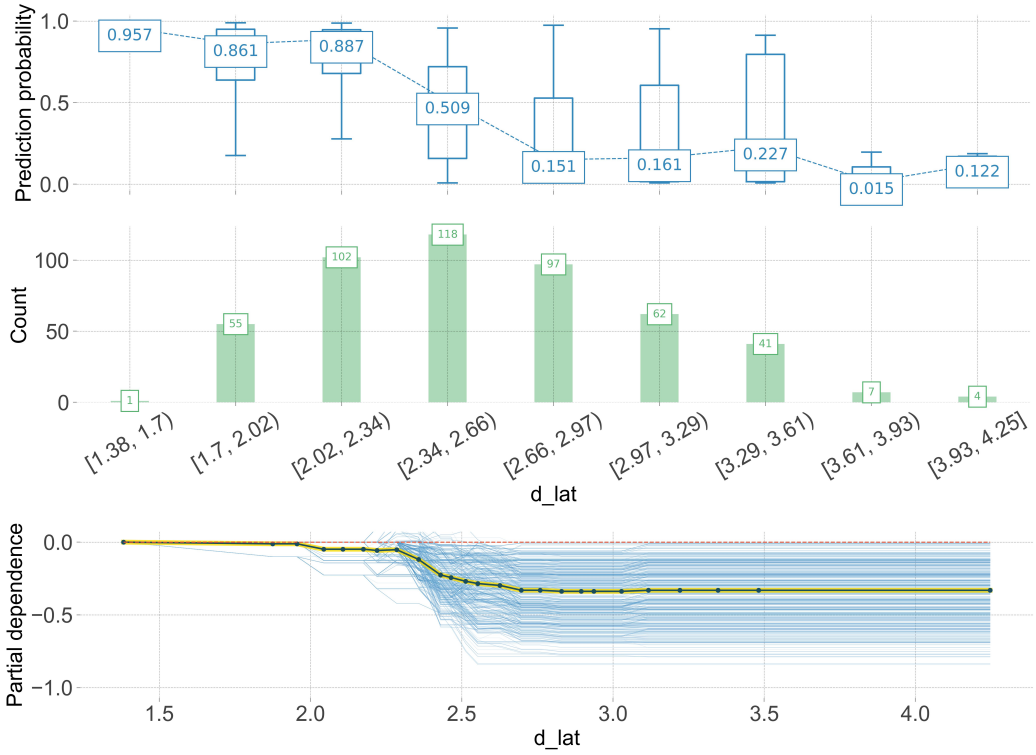


(a) Longitudinal distance

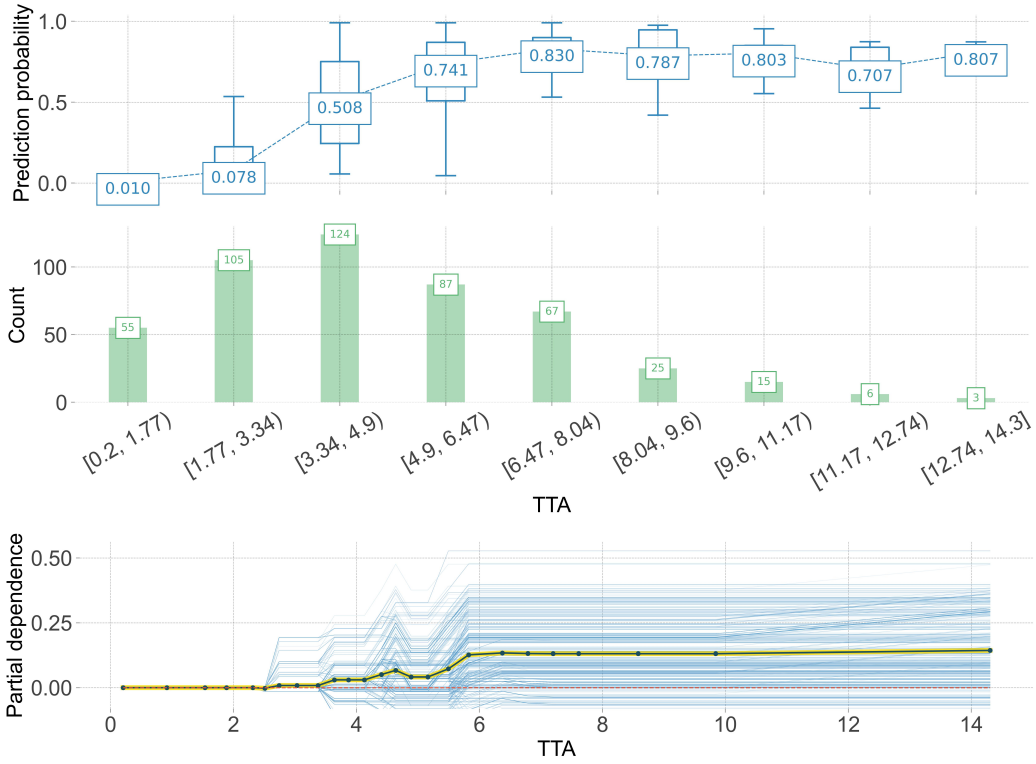


(b) Vehicle's acceleration

Figure 3.14: PDPs for the longitudinal distance and the vehicle's acceleration in the pedestrian's decision-making process.

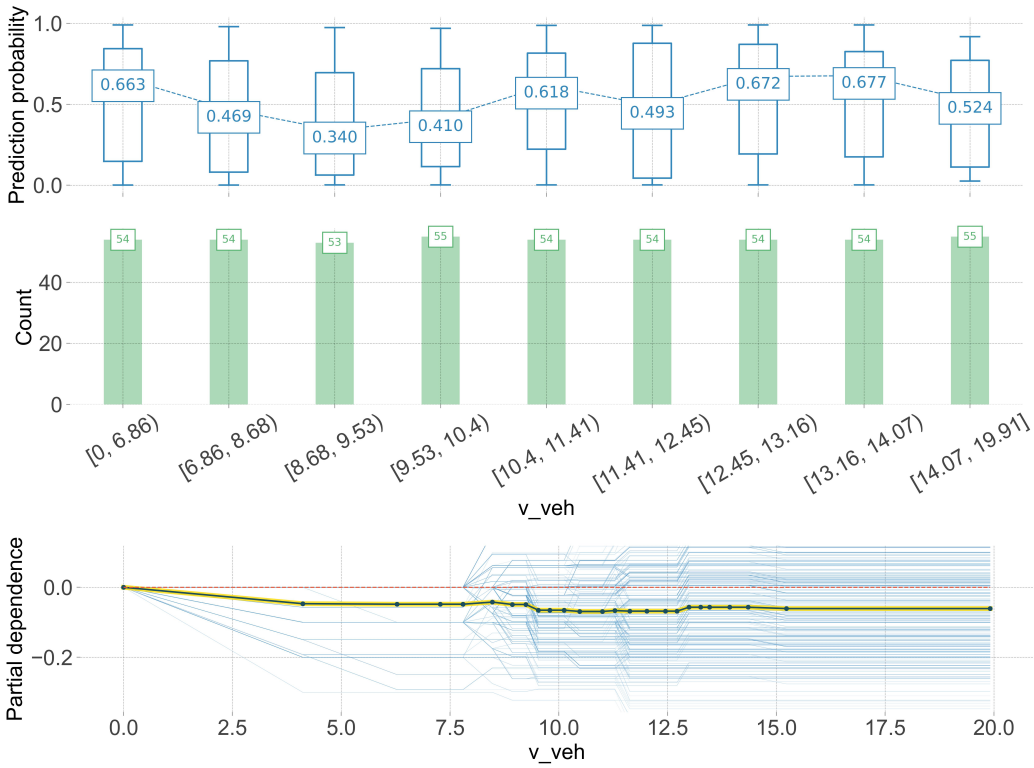


(a) Lateral distance



(b) TTA

Figure 3.15: PDPs for the lateral distance and the TTA in the pedestrian's decision-making process.



(a) Vehicle's speed

Figure 3.16: PDPs for the vehicle's speed in the pedestrian's decision-making process.

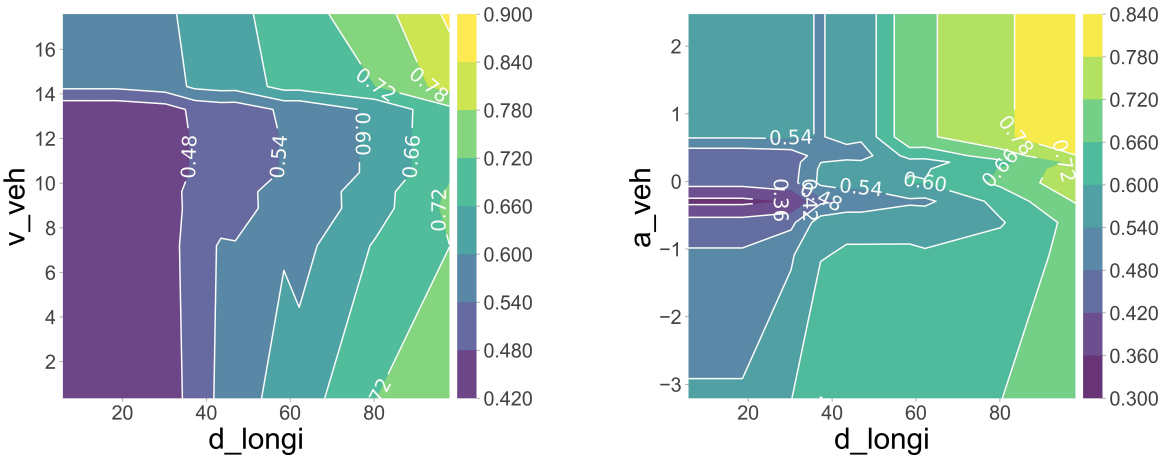
versely, when the vehicle decelerates slightly with a smaller longitudinal distance, the likelihood of pedestrians crossing the road reaches its lowest. This is attributed to the behavior of some drivers (27%) who lightly engage the brake pedal upon spotting the pedestrian. Consequently, pedestrians may perceive the safety risk of crossing the road as too high with a short longitudinal distance, prompting them to yield to the approaching vehicle.

3.4.3 Driver-Pedestrian Interaction

Since pedestrians' speed and longitudinal distance are the most significant variables impacting driver and pedestrian decision-making respectively, we have formulated the following definitions to further explore the relationship between these parameters and the decision-making process:

$$\tan\alpha = \frac{\mathrm{v}_{\mathrm{veh}}}{\mathrm{v}_{\mathrm{ped}}} \tag{3.9}$$

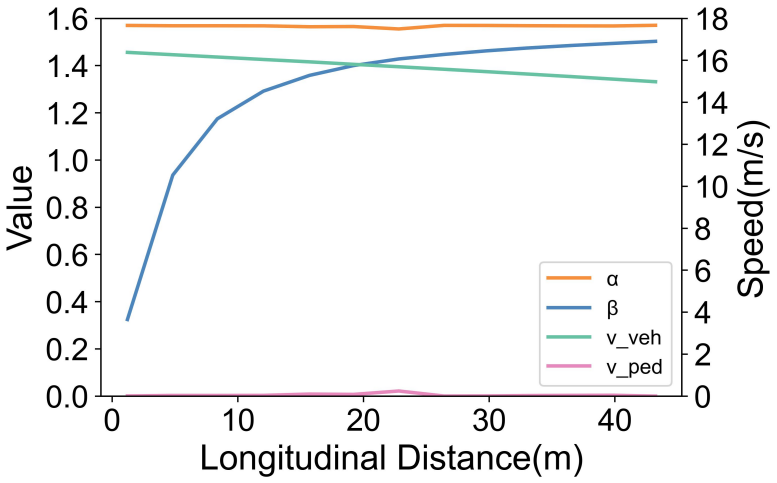
$$\tan\beta = \frac{\mathrm{d}_{\mathrm{longi}}}{\mathrm{d}_{\mathrm{lat}}} \tag{3.10}$$



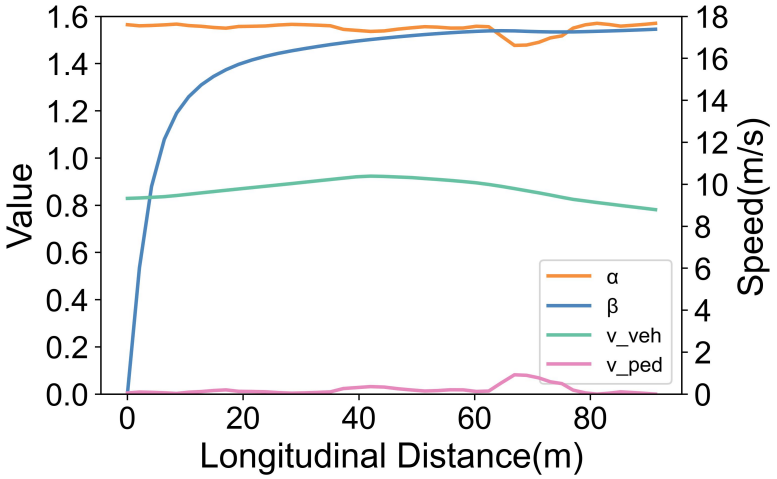
(a) (b)

Figure 3.17: The mutual impact between longitudinal distance and vehicle speed as well as vehicle acceleration respectively.

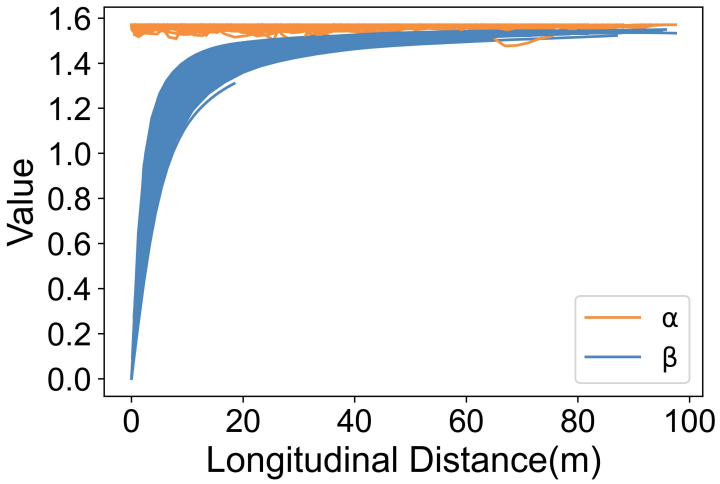
On the one hand, in situations where a pedestrian tends to yield, Fig. 3.18 illustrates the relationship between α and β. The graph distinctly shows a continuous trend where α remains consistently greater than β if a pedestrian allows a vehicle to proceed first. Fig. 3.18 also provides two illustrative examples. It is evident, particularly in Fig. 3.18a, that the pedestrians' speed stays essentially constant, hovering around 0. As the ratio of vehicle speed to pedestrian speed tends to infinity when the pedestrian speed approaches zero, α approaches a value indefinitely close to $\pi/2$. Given that the longitudinal distance decreases while the lateral distance remains constant and the arctan function becomes monotonically growing, the value of β drops.



(a)

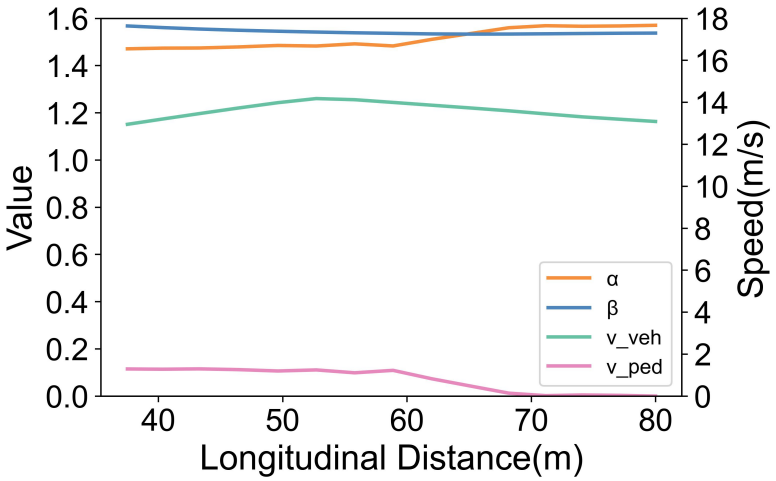


(b)

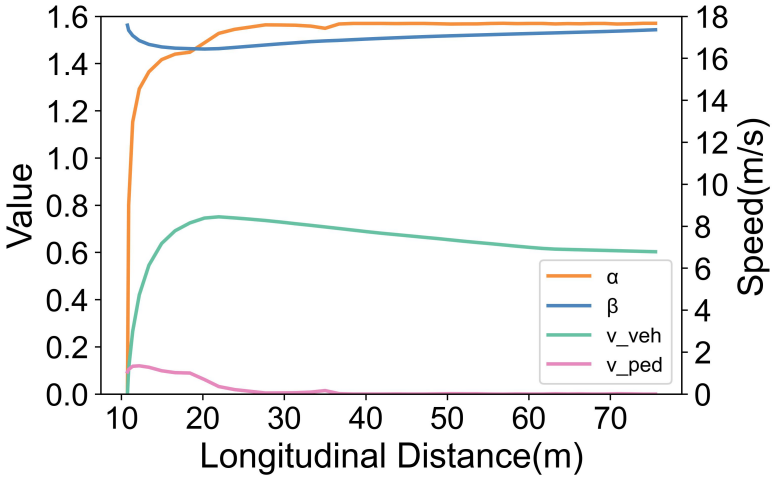


(c)

Figure 3.18: The relationships between α and β when the pedestrians yield. (a) A normal case of the relationships between α and β. (b) A special case of the relationships between α and β. (c) Overview of the relationships between α and β in all cases.



(a)



(b)

Figure 3.19: Two examples of relationships between α and β when the vehicle yields.

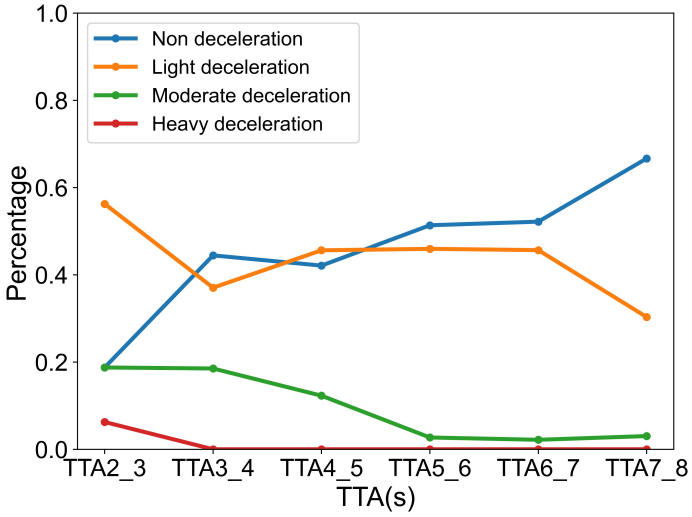
In contrast, Fig. 3.18b presents a special scenario where α is not always greater than β. This situation arises because the pedestrian does not make a one-time decision and may modify it during the interaction process. Specifically, in this scenario, although α initially holds a higher value than β, the pedestrian begins to accelerate, causing α to decline at a longitudinal distance of 77 m. The value of α equalizes with β when the pedestrian's speed reaches 0.35 m/s. Subsequently, α consistently maintains a lower than β. The pedestrian's speed peaks at 0.92 m/s when the longitudinal distance is 66.8m. However, post this peak, the pedestrian's speed begins to decline. At a longitudinal distance of 63.3m, the pedestrian's speed reduces to 0.35 m/s, aligning α equal to β again before gradually exceeding β. This observation implies that the pedestrian initially intends to

cross the road while the longitudinal distance falls within the range of 63.3 m to 77 m. However, the pedestrian ultimately changes his decision, opting to yield to the vehicle instead. Therefore, we can infer that the pedestrian yields to the vehicle when α exceeds β.

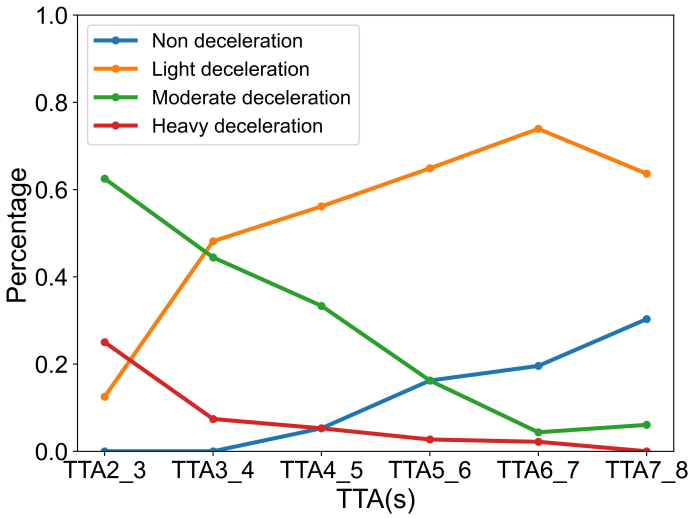
On the other hand, Fig. 3.19 presents two instances depicting the relationship between α and β in scenarios where the pedestrian opts to cross the road. Both figures clearly indicate that if the pedestrian eventually crosses the road before the vehicle, the value of α is initially greater than β and then decreases to be lower than β. Specifically, in Fig. 3.19a, where the pedestrian's speed is nearly 0 and remains stable, the pedestrian is likely in the decision-making process, showing no definitive inclination to cross. During this period, α exceeds β, aligning with our earlier conclusion. However, when the longitudinal distance reaches 68.3 m, the pedestrian's speed begins to notably increase, indicating the decision to cross. Although the vehicle's speed continues to rise, the change in pedestrian speed significantly influences the value of α, causing α to decrease and eventually become smaller than β.

Similarly, Fig. 3.19b demonstrates a scenario where the pedestrian initially moves at a slow pace as the longitudinal distance decreases from 75 to 27 meters. The pedestrian accelerates when the distance falls below 27 meters. Due to the close longitudinal distance, when the driver perceives that the pedestrian is about to cross the road, he engages the brake pedal. Consequently, the value of α drops below β, rapidly approaching zero. Hence, it can be inferred that if α is less than β, the pedestrian chooses to cross the road.

In summary, we can draw the conclusion that if α exceeds β, the pedestrian is likely to yield to the vehicle. Conversely, if α is less than β, the pedestrian is inclined to cross the road before the approaching vehicle. This finding holds significant value in dynamically predicting pedestrian decisions using easily accessible variables.



(a)



(b)

Figure 3.20: The proportion of drivers' behavior under different TTA conditions.

3.4.4 Driver's Yielding Behavior Analysis

This part investigates the actual drivers' behavior when yielding to pedestrians. As indicated in the preceding section, when the value of α exceeds β in all scenarios where vehicles yield to pedestrians, the pedestrian exhibits no obvious intention to cross the road. Conversely, when α is less than β, the pedestrian displays a clear intention to cross. Therefore, this study aims to investigate the driver's behavior in all vehicle-yielding situations when the pedestrian's intention to cross the road is either ambiguous or evident. The primary focus of this study is on driver behavior when the TTA ranges between 2 and 8 seconds. Using cluster analysis [208], [209], the driver behavior is categorized into

four groups based on vehicle acceleration data: non-deceleration (greater than 0), light deceleration (0, -1.90 m/s^2), moderate deceleration (-1.90 m/s^2, -5.26 m/s^2), and heavy deceleration (-5.30 m/s^2, -10.15 m/s^2). Fig. 3.20 illustrates observed driver behaviors in the yielding scenarios.

Fig. 3.20a displays driver behavior when α exceeds β under varying TTA conditions. In instances where the TTA ranges from 2s to 3s, most drivers tend to lightly brake when they are uncertain about the pedestrian's intention. Approximately 20% of drivers do not engage in any braking, while an equivalent percentage opts for moderate braking. The minority of drivers opt for heavy braking. However, when the TTA exceeds 3s, no driver chooses to brake heavily. The proportion of moderate braking similarly drops to 2% at the TTA from 5s to 6s and remains relatively constant thereafter. The percentage of drivers who do not decelerate increases, reaching its peak at 66% between 7s and 8s. In contrast to scenarios where there is no deceleration, light deceleration exhibits a slightly diminishing trend.

In Fig. 3.20b, the driving behavior of drivers is observed when α is less than β across various TTA scenarios. When the TTA ranges from 2s to 3s, the majority of drivers opted for moderate braking upon understanding the pedestrian's intent to cross, followed by heavy braking(25%) and light braking(12.5%) respectively. Notably, none of the drivers chose to accelerate through the crossing point. The percentages of moderate and heavy braking, however, decline as the TTA rises. There is no instance of significant braking observed between 7s and 8s. Instead, an increase in the proportion of no deceleration and light braking is witnessed. Specifically, the proportion of light braking rises rapidly during the period from 2s to 4s and remains the highest throughout. Concurrently, the proportion of no deceleration gradually increases, surpassing that of heavy and moderate deceleration, reaching the second highest and accounting for 30% eventually.

Based on the aforementioned analysis, we can draw the following conclusions: Firstly, when a pedestrian is within the driver's line of sight but their intentions are uncertain, and the TTA falls between 2 and 5 seconds, most drivers tend to slightly reduce their speed. However, if the TTA spans 5 and 8 seconds, over half of the drivers do not slow down. Secondly, when the driver notices that the pedestrian intends to cross the street, they usually brake moderately or heavily when the TTA ranges from 2 and 3 seconds. In cases where the TTA surpasses 3 seconds, more than half of drivers tend to only slightly decrease their speed.

3.5 Chapter Conclusion

This chapter begins by highlighting limitations in existing research on interactive behaviors between human drivers and pedestrians. First, many prior studies focus on identifying variables that explain pedestrian crossing behavior rather than predicting pedestrian crossing intention. Second, the variables often employed to model human decision-making are difficult to obtain or estimate in real-world scenarios, limiting their applicability in practical decision-making models. To address these gaps, this study investigates the vehicle-pedestrian interaction at unsignalized intersections using the AdaBoost method in a virtual reality environment. In contrast to existing research, our study focuses solely on kinematic parameters in the analysis of interactions and has conducted a comprehensive analysis of all kinematic parameters, providing valuable insights into vehicle-pedestrian interaction dynamics at unsignalized intersections. The analysis demonstrates that a pedestrian's decision is primarily influenced by the longitudinal distance and the vehicle's acceleration, while the pedestrian's own speed and longitudinal distance also significantly impact whether the approaching vehicle will yield. Based on these findings, a simple mathematical relationship between kinematic parameters and pedestrian intentions was established, providing a practical tool for dynamically inferring crossing intentions during interactions.

In exploring scenarios where vehicles yield to pedestrians, the analysis uncovers interesting patterns, which contribute insights to our understanding of real-world interactions at unsignalized intersections and lay a foundation for improving the development of autonomous vehicle technology. When it is uncertain whether a pedestrian intends to cross, the majority of drivers only slightly reduce their speed even in lower TTA cases. Surprisingly, over half of the drivers do not decelerate at all with higher TTA. Conversely, when the pedestrian's crossing intention is clear, a substantial proportion of drivers opt for moderate or heavy deceleration, particularly in low TTA situations. If the TTA exceeds 5 seconds, light deceleration becomes the most common choice.

Studying the human driver-pedestrian interaction allows a deeper understand of how human drivers make decisions, which in turn informs the development of a human-like AV decision-making model. By emulating human drivers' decision patterns, the AV can behave in ways that pedestrians can readily interpret and predict, thereby enhancing its social acceptance. In the following chapters, a decision-making model for AV-pedestrian interaction is developed using the insights gained from this analysis. Specifically, the significant factors influencing driver decision-making are integrated into the AV model through the design of the reward function and the selection of state variables, while the key determinants of pedestrian decision-making are incorporated into the construction

of the pedestrian behavior model. The next chapter focuses on an interaction scenario between a single AV and a single pedestrian, while the subsequent chapter extends the scenario to address more complex situations involving multiple pedestrians. These models aim to enable the AV to plan and control its movements in a natural and human-like manner when navigating complex urban areas.

Although this work studies human-driven vehicle and pedestrian behavior patterns to provide insights for AV-pedestrian interaction modeling. However, pedestrians may respond differently when interacting with AVs compared to human-driven vehicles. Therefore, using pedestrian behavior observed in human-driver scenarios to model AV-pedestrian interactions may introduce biases, which is acknowledged as a limitation of the current study. As a potential extension, future work could incorporate dedicated AV-pedestrian interaction datasets to more explicitly and accurately capture the pedestrian's behavior. In addition, the current VR experiment is restricted to interactions with a single pedestrian due to the limitations of the experimental equipment. Future work will incorporate multi-pedestrian scenarios to capture more diverse and realistic interaction patterns.

Chapter 4

Dynamic Game-Theoretical Decision-Making Framework for AV in Single-Pedestrian Interactions at Unsignalized Intersections

4.1 Background

This chapter focuses on AV decision-making in pedestrian interaction scenarios at unsignalized intersections, where potential conflicts may occur, as shown in Fig. 1.1a. The objective is to model the dynamics of AV-pedestrian interactions during crossing situations and to develop an explainable and trustworthy decision-making framework for AVs operating in such environments.

Previous studies have investigated vehicle-pedestrian interactions, often relying on statistical methods [11], [176] or describing interactions as one-time events [210]. However, pedestrian behavior, characterized by unpredictability and dynamism, presents challenges for such approaches. Their movements can quickly change [211], introducing uncertainty into interactions that traditional methods struggle to capture. By considering uncertainty and dynamic interactions, the POMDP framework [212] provides a modeling approach for decision-making challenges that closely mirror real-world conditions. In the context of self-driving decision problems, POMDP is commonly employed to capture the incomplete observability and uncertainty in the AV's surrounding environment. The study [213] regarded pedestrians' target position as an unobservable variable in the POMDP model to capture the decision-making and planning behavior of autonomous vehicles navigating among many pedestrians. Similarly, an AV-pedestrian interaction model was proposed in

[214] to address complex decision-making challenges arising from the uncertain crossing intention of pedestrians in urban environments by leveraging the POMDP framework. Although a few studies have applied the POMDP framework to vehicle-pedestrian interactions, its potential in this domain remains underexplored.

Game theory is frequently used to model the interaction between vehicles and pedestrians. However, most studies assume that all players follow the Nash equilibrium [215], [216], possessing unlimited computational reasoning ability to compute optimal actions and perfect rationality to execute them, thus maximizing their utility function in decision-making. In reality, individuals often deviate from the Nash equilibrium due to cognitive limitations [217]. Constrained by limited time and computational resources, they struggle to engage in extensive or complex reasoning, thereby failing to consistently calculate optimal actions or avoid errors in challenging scenarios. Hence, considering human reasoning levels and bounded rationality is essential to develop more accurate models of real-world behaviors.

To address these limitations, our study proposes a novel framework that combines POMDP and behavioral game theory to tackle the AV decision-making problem within complex and dynamic environments. Fig. 4.1 shows our proposed framework for AV-pedestrian interaction at an unsignalized intersection. In this work, we employ the POMDP framework to dynamically model the decision-making process of the AV in an environment with incomplete information and uncertainty. Furthermore, we use a behavioral game theory model to describe AV-pedestrian interaction, both the AV and pedestrian modeled as DB-QCH models. At each time step, the AV model updates its beliefs about its opponent's reasoning level and rationality based on extended Bayesian Estimation. A trained neural network calculates the predicted optimal action, which is then translated to an action space using a Gaussian distribution function. An iterative reasoning model is established to deduce the optimal strategies for both oneself and the opponent at each level, computed via the Monte Carlo Tree Search (MCTS) method. While the pedestrian model is also constructed as a DB-QCH model, their action space remains fixed over time, which differs from the AV model.

To enhance humans' understanding of the interaction process and resolve intersection conflicts in a human-involved interactive environment, this work makes the following main contributions: 1). The POMDP framework and behavioral game theory are integrated to address the uncertainty and dynamic interaction between the AV and the pedestrian. 2). To accurately capture the decision-making processes, both the AV and pedestrian are modeled as DB-QCH models. This modeling approach provides a comprehensive understanding of interaction dynamics and facilitates more realistic simulations. 3). A trained neural network based on data from our previous experiments is developed to guide MCTS

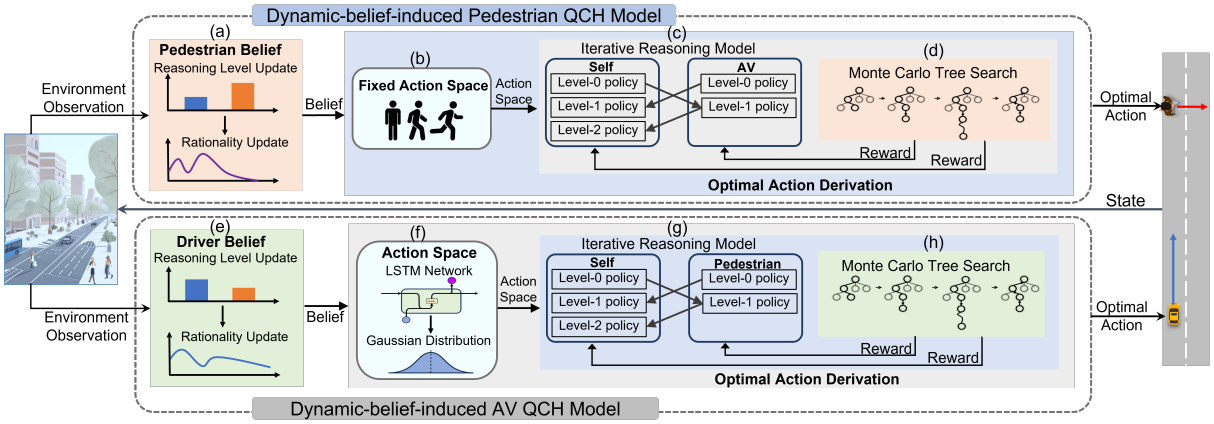


Figure 4.1: The proposed framework of AV-pedestrian interaction at the unsignalized intersection. Subparts (a) and (e) represent the belief update process, adjusting reasoning levels and rationality; Subparts (b) and (f) illustrate the action space generation; Subparts (c) and (g) depict the reasoning mechanisms of predicting and adapting to the opponent's strategies for decision optimization; Subparts (d) and (h) show the Monte Carlo Tree Search for action evaluation through reward calculations.

in exploring the continuous action space of AV, thereby facilitating effective and efficient decision-making. 4). This work introduces variables to quantify human bounded rationality and is the first to propose a dynamic updating mechanism for rational values based on the observed environment, enabling adaptive decision-making by AVs in real-time. These concentrated efforts pave the way for an explainable and trustworthy AV decision-making system, leading to safer and more efficient navigation of AVs in such an interactive environment where humans are involved.

4.2 Problem Formulation

This work focuses on addressing the challenge posed by conflicts arising at the unsignalized intersection, where both the pedestrian and the AV intend to cross simultaneously. Specifically, it aims to develop continuous decision-making strategies for AVs navigating safely and efficiently through such a scenario. As the AV lacks knowledge about the opponent's intelligence level and rationality in a dynamic and interactive environment, we model the interaction between the AV and the pedestrian using a POMDP framework. The model is defined by the tuple:

$$\langle N, S, A, T, O, J, B \rangle$$

- $N = \{0, 1\}$: Represents the two players, where 0 denotes the AV and 1 denotes the pedestrian.

- S: A finite set of states, where $s_t \in S$ signifies the state of the environment at discrete time step t. The state s_t includes the position and velocity of both the AV and the pedestrian, as well as the acceleration of the AV.
- $A = \{A^0, A^1\}$: Defines the action space, with A^0 representing the AV's actions and A^1 denoting the pedestrian's actions.
- T: The state transition dynamics, expressed as $s_{t+1} = T(s_t, a_t^0, a_t^1)$ for an action pair $(a^0, a^1) \in A$. This function describes how the environment transitions from one state to another based on the actions of both players.
- $O = \{O^0, O^1\}$: Represents the partially observable state. We assume that each agent's action can be observed, along with certain physical information (e.g., speed, acceleration, distance), while implicit information (e.g., reasoning level, rationality degree) remain unobservable.
- $J = \{J^0, J^1\}$: The utility function for each agent. The utility $J_t^i = J^i(s_t, a_t^0, a_t^1)$, $i \in N$, depends on both the agent's action and the opponent's action.
- $B = \{B^0, B^1\}$: The belief in the opponent's intelligence level and rationality, with $b_t^0 \in B^0$ and $b_t^1 \in B^1$.

For the AV model, the goal is to determine a sequence of optimal actions. The optimization problem can thus be formulated as follows:

$$\begin{aligned} \underset{\pi}{\text{maximize}} \quad & \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma_t J_t^0(s_t, a_t^0, a_t^1, b_t^0) \mid a_t^0 \sim \pi\right] \\ \text{subject to} \quad & s_{t+1} = T(s_t, a_t^0, a_t^1), \\ & b_{t+1}^0 = \rho(b_t^0, o_t^0), \\ & a_t^0 \in A^0, a_t^1 \in A^1, \\ & o_t^0 \in O^0, b_t^0 \in B^0 \end{aligned} \tag{4.1}$$

where γ represents the discount factor within the range of $(0, 1]$, while ρ denotes the belief update function. And s_t represents the state of the environment at the time step t, while a_t^0 and a_t^1 represents the actions of AV and pedestrian at time t, respectively.

4.3 Methodology

This section provides a detailed description of the approaches we use to model the interaction between the AV and the pedestrian at the unsignalized intersection and solve the above problem.

4.3.1 Action Space Generation

Consider a scenario where both a pedestrian and an AV are moving toward an intersection. To avoid collision, they have to select an appropriate action. For example, the pedestrian can choose to cross the road or wait, while the AV can either slow down or maintain its speed. To make an informed decision, each agent model first defines its own action space.

For our AV model, the dynamic decision-making process aims to produce a sequence of expected accelerations. However, in actual scenarios, the AV's acceleration range exists in a continuous space, posing challenges for methods like MCTS, which typically excel in discrete action spaces. To address this, we employ a pre-trained neural network model to guide MCTS through the continuous action space of AVs in Fig. 4.1f.

We have opted for the LSTM network as our neural network model. LSTM, a subtype of RNN, is good at processing and predicting time series data, adeptly capturing temporal dependencies [218]. Unlike traditional RNNs, LSTM overcomes long-term dependency issues through its gate mechanisms(including input gate, forget gate, and output gate), effectively retaining and leveraging long-term information [219]. This gives LSTM a significant advantage in handling complex time series data in autonomous driving scenarios.

Acceleration prediction in autonomous driving presents a highly temporal problem, as a vehicle's acceleration depends not only on its current state but also on past states and actions. Traditional RNNs often encounter problems like gradient vanishing or exploding when dealing with long-term dependencies [220], making it difficult to effectively capture long-term dependency information. LSTM, with its unique gating mechanism, can maintain and transmit key information across lengthy time series, avoiding these shortcomings of traditional RNNs [221].

Given the temporal dynamics and complexity of acceleration prediction tasks, we choose LSTM as our preferred model for anticipating AV acceleration. Its ability to use historical data enhances prediction accuracy and stability, providing robust support for our decision-making system.

Training data for the LSTM model is sourced from our prior vehicle-pedestrian interaction experiments [222], conducted using VR technology. This experiment yielded dynamic interaction data, including the absolute positions of pedestrians and vehicles, vehicle speeds, and driver inputs like steering, throttle, and brakes. Through data processing, we extracted relevant variables such as vehicle speed, acceleration, relative distances, time-to-arrivals, pedestrian speeds, and vehicle yielding status at each time step for every scenario.

This data underwent training in the LSTM model, which, post-training, can ingest state information at each time step and output the corresponding anticipated acceleration. This acceleration is treated as the mean of a Gaussian distribution, from which N accelerations are sampled, yielding $N+1$ possible accelerations. Subsequently, MCTS is employed to explore these $N+1$ actions, leveraging this neural network model's output as an initial guide to enhance MCTS's efficiency in navigating the continuous action space. Through this integration of the neural network and MCTS methods, we can improve the decision-making ability of autonomous vehicles in complex dynamic interactive environments.

In contrast, within the pedestrian model, the action space pertains to the pedestrian's speed. Since pedestrian speed can change rapidly, using a neural network model with Gaussian distribution sampling for action spaces, as used in the AV model, is less effective. Instead, we adopt a discrete action selection space to simplify our model in order to improve computational efficiency, as shown in Fig. 4.1b. Starting from 0 m/s, we discretize the speed at 0.1 m/s intervals. This method is simpler than the one used for the AV model and adequately meets our needs. Unlike the AV model, where elements in the action space dynamically change, the pedestrian's action space remains fixed at each time step.

4.3.2 Dynamic-belief-induced Quantal Cognitive Hierarchy Model

4.3.2.1 Quantal Cognitive Hierarchy model

The Quantal Cognitive Hierarchy model is a behavioral game theory model used to describe the behavior of bounded rational individuals in games. It integrates the quantal response (QR) model into the traditional cognitive hierarchy (CH) model.

In the CH model, agents are characterized by different cognitive levels, each associated with varying degrees of reasoning abilities and consideration of others' behavior. Higher levels indicate greater reasoning capabilities and more consideration of opponents' actions. At each level, agents simulate their opponents' behavior under the assumption

that opponents operate at lower levels. Each agent's cognitive level is denoted by k (where $k = 0, 1, 2$). Level-0 agents are regarded as non-strategic, generating their strategies independently and without considering opponents' behavior, often through uniform random selection or simple heuristic methods. Conversely, strategic agents at level-k (where $k > 0$) engage in a more sophisticated decision-making process. They assume their opponents operate at level-j, where $j < k$, and respond accordingly with optimal strategies. As human cognition generally operates at levels-0, 1 or 2, with level-3 or higher rarely observed [223], this paper considers the maximum value of level-k to be 2 in order to capture realistic interactions and reduce computational complexity.

The QR model introduces the concept of bounded rationality, where agents do not always choose the optimal strategy but select strategies with certain probabilities based on expected payoffs when making decisions. In this model, bounded rationality is represented by the parameter λ (where $\lambda \in [0, \infty)$), which measures the degree of rationality. A higher λ indicates more rational behavior, while a lower value reflects greater randomness in decision-making. The probability $P(a_i)$ that an agent i chooses a particular strategy a_i given the opponent's action is described by the quantal response function:

$$P(a_i) = \frac{e^{\lambda Q(a_i, a_{-i})}}{\sum_{a'_i \in A} e^{\lambda Q(a'_i, a_{-i})}} \tag{4.2}$$

where $Q_i(a_i, a_{-i})$ is the expected payoff for agent i when choosing strategy a.

Equation 4.2 shows that the probability of selecting a strategy increases with its expected payoff, meaning individuals are more likely to select a strategy with a higher expected payoff but may also opt for those with lower returns. As λ approaches infinity, the model approximates perfect rationality, where the highest payoff strategy is always chosen. Conversely, when λ is close to zero, the choice of strategy becomes completely random.

By combining ideas from the CH model and QR model, the QCH model offers insights into how individuals probabilistically select strategies at different cognitive levels, thereby enhancing our understanding of bounded rational behavior. In our study, we adopt the QCH model to represent the decision-making process for both the AV and the pedestrian. This model captures the varying levels of intelligence k and rationality λ of each opponent, which are unobservable to each other. Table 4.1 summarizes these two key parameters. Here, we detailed the decision-making process specifically using the AV QCH model as an example.

Table 4.1: Summary of the two key parameters.

Parameter	Notation	Operationalization	Interpretation
Reasoning level	k	$k = 0, 1, 2, \ldots$ introduces recursive reasoning to allow agents to consider others' behavior, where higher levels account for more sophisticated reasoning.	Simulates the reasoning process behind behavioral complexity, enabling the agent model to predict and adapt to opponents' strategies.
Bounded rationality	λ	$\lambda \in [0, \infty)$, introduces randomness to cause the agent to deviate from optimality, with lower values indicating more randomness in decision-making.	Models deviations from purely optimal behavior, making the agent model more realistic.

At each level-k, the AV evaluates its potential actions by calculating the expected payoff Q for each action given the current state s_t. This evaluation also considers the pedestrian's policy from the preceding level-$(k-1)$. The AV then makes a quantal best response to the pedestrian's level-$(k-1)$ policy. In addition, the AV model reasons the potential actions available to the pedestrian at this level, preparing for subsequent policy computations. As illustrated in subparts (c) and (g) of Fig. 4.1, the policy at each level for the AV and its opponent is developed through a sequential and iterative process, starting from level-0 to higher levels. For example, if a level-0 pedestrian decides to cross the road, a level-1 AV will consider the pedestrian's action and choose to yield. In turn, a level-2 pedestrian will observe the level-1 AV's yielding behavior and choose to cross the road. Similarly, if a level-0 AV chooses to cross the road, a level-1 pedestrian will yield after considering the AV's action, and a level-2 AV will then choose to cross the road based on the pedestrian's yielding behavior. In our study, we assume that the level-0 agent lacks understanding of pedestrian intentions or higher-level policies, and instead treats pedestrians as stationary obstacles to compute its actions. In contrast, the level-k (where $k > 1$) agent regards its opponents as level-$(k-1)$ agents. Specifically, the quantal response function is used to compute the policy:

$$\pi^{i,k,\lambda^i}(a^i_j) = \frac{e^{\lambda^i Q^k(s_t, a^i_j, \pi^{-i,k-1,\lambda^{-i}})}}{\sum_{a' \in A^i} e^{\lambda^i Q^k(s_t, a', \pi^{-i,k-1,\lambda^{-i}})}} \tag{4.3}$$

After computing strategies for all levels, we can derive the optimal strategy using initial beliefs. Finally, we can determine the optimal action for the AV, selecting the action associated with the highest mixed strategy value. The algorithm for this process is shown in Algorithm 4.

Introducing the QCH model provides the advantage of simultaneously accounting for the agent's limited rationality and reasoning level, thus making the model more realistic. Through this approach, we can more accurately simulate the decision-making behaviors of real-world agents and gain insights into the interactions between AVs and pedestrians.

Algorithm 2: QCH Model Iterative Reasoning to Compute Optimal Action.

Input: N: Player set, A: Possible action set, s_t: Current state, $\pi^{i,0}$: The level-0 policy for agent i, K: Maximum cognitive level, b_k: Belief about the opponent's level, λ: Rationality degree

Output: Optimal action

Initialize $agent\_policy \leftarrow []$;
Initialize $mix\_policy \leftarrow []$;
Append $\pi^{i,0}$ to $agent\_policy$;
for $k = 1$ **to** K **do**
  for *each player* $i \in N$ **do**
    for *each action* $a^i_j \in A^i$ **do**
      Compute the payoff $Q^{i,k}(s_t, a^i_j, \pi^{-i,k-1,\lambda^{-i}})$;
    Compute the policy $\pi^{i,k,\lambda^i}(a^i_j) = \frac{e^{\lambda^i Q^{i,k}(s_t, a^i_j, \pi^{-i,k-1,\lambda^{-i}})}}{\sum_{a' \in A^i} e^{\lambda^i Q^{i,k}(s_t, a', \pi^{-i,k-1,\lambda^{-i}})}}$;
    Append π^{i,k,λ^i} to $agent\_policy$;
for $k = 1$ **to** K **do**
  $mix\_policy \leftarrow mix\_policy + b_k[k-1] \cdot agent\_policy[k]$;
$optimal\_index \leftarrow \arg\max(mix\_policy)$;
$optimal\_action \leftarrow A[optimal\_index]$;
return $optimal\_action$;

4.3.2.2 Dynamic Belief Update

For AVs, the opponent's reasoning level-k and rationality degree λ are not directly observable. Pedestrian behavior is dynamic and may constantly change. If AVs always use fixed values of pedestrian's cognitive states for best response calculation during interactions with pedestrians, they will be unable to effectively identify and adapt to changes in pedestrian behavior. This would make AVs appear less intelligent and human-like, as humans can quickly recognize and respond to sudden behaviors. The Bayesian approach allows AVs to continuously learn and update their beliefs about pedestrians' reasoning level-k and rationality λ during interactions(Fig. 4.1a, Fig. 4.1e).

At time step $t = 0$, the agent i establishes an initial belief $b_{k,0}$ about the pedestrian's reasoning level, according to the initial environmental state and our prior experimental data on human-vehicle interactions. Throughout the game reasoning process, its QCH model iteratively predicts the expected utility of the opponent's potential actions across each reasoning level k for the next state s_{t+1}, alongside computing the associated probability $P(s_{t+1}, a^{-i}_{t+1}|k)$, where $s_{t+1} \in S, a^{-i}_{t+1} \in A^{-i}, k \in \mathbb{N}$ for each action. Upon observing the opponent's latest action a^{-i}_{t+1} at time step $t+1$, the agent model identifies the probability value of the action that is closest to the actual observed action at each reasoning level k. Subsequently, followed by normalizing the probabilities to ensure coherence, it updates its

belief $b_{k,t+1}$ concerning its opponent's reasoning level across all levels using the Bayesian equation:

$$P(k|s_{t+1}, a_{t+1}^{-i}) = \frac{P(s_{t+1}, a_{t+1}^{-i}|k) b_{k,t}(k)}{\sum_{k' \in \Theta} P(s_{t+1}, a_{t+1}^{-i}|k') b_{k,t}(k')} \tag{4.4}$$

where $P(k|s_{t+1}, a_{t+1}^{-i}) \in b_{k,t+1}$, and Θ represents all possible values for the reasoning level.

Our model differs from others by dynamically updating the belief about the opponent's rationality degree λ, rather than relying on constant values. The Bayesian approach is employed to capture changes in the agent i's understanding of its opponent in this parameter. Initially, agent i has prior knowledge regarding the distribution of λ, denoted as $f_t(\lambda)$, indicating its estimation of the opponent j's rationality at the current time step t. Since $\lambda \in [0, \infty)$, it is treated as a continuous variable. Therefore, we use a Bayesian updating method suitable for continuous variables [224]:

$$f_{t+1}(\lambda|a_t^j) = \frac{P(a_t^j|\lambda) f_t(\lambda)}{\int_0^\infty P(a_t^j|\lambda')|f_t(\lambda')d\lambda'} \tag{4.5}$$

Considering the varying reasoning level of agent j, the actions it takes will correspond to its specific reasoning level. Consequently, we can extend the Bayesian formula accordingly:

$$f_{t+1}(\lambda|a_t^j, k) = \frac{P(a_t^j|k, \lambda) f_t(\lambda)}{\int_0^\infty P(a_t^j|k, \lambda')|f_t(\lambda')d\lambda'} \tag{4.6}$$

Since it involves integrating the function value over an infinite interval, it is challenging to compute Equation 4.6 after multiple iterations. The conjugate prior distribution proves highly effective in addressing this issue [225]. In Bayesian methodology, if the posterior distribution belongs to the same family as the prior distribution, the prior distribution is referred to as a conjugate prior distribution. The benefit of using a conjugate prior lies in its ability to simplify the Bayesian update process, requiring only the adjustment of parameters within the conjugate prior distribution to complete the Bayesian inference update. Our work considers the following family of distributions [226]:

$$f(\lambda; Q, n_0, n_1, \ldots, n_K) = \frac{e^{\lambda Q} / \prod_{k=0}^{K} (\sum_{l=1}^{m} e^{\lambda Q_{a_{l,k}}})^{n_k}}{\int_0^\infty e^{\lambda' Q} / \prod_{k=0}^{K} (\sum_{l=1}^{m} e^{\lambda' Q_{a_{l,k}}})^{n_k} d\lambda'} \tag{4.7}$$

where $n_k \in \mathbb{N}, \forall k = 0, 1, 2, \ldots, K$, representing the number of occurrence agent j's reasoning level corresponds to k. $f(\lambda; Q, n_0, n_1, \ldots, n_K)$ is a probability density function as $\int_0^\infty f(\lambda; Q, n_0, n_1, \ldots, n_K) d\lambda = 1$.

At each time step t, through game reasoning, agent i calculates the expected utility $Q_{a_{j,k}}$ for each action that their opponent may take at each level and predicts the reasoning level k of agent j. By using the prior distribution $f(\lambda;Q,n_0,n_1,\ldots,n_K)$ and the observed action a_j taken by its opponent, agent i can then update the distribution of λ, $f(\lambda;Q+Q_{a_{j,k}},n_0,n_1,\ldots,n_k+1,\ldots,n_K)$, and use it as the prior distribution at the next time step $t+1$. It is known that when employing Equation 4.7 as the prior distribution, updating the parameters Q and n_k to obtain the posterior distribution $f(\lambda;Q+Q_{a_{j,k}},n_0,n_1,\ldots,n_k+1,\ldots,n_K)$ can be directly achieved without the need for intricate integral calculations.

Theorem 1: Given the prior distribution $f_t(\lambda;Q,n_0,n_1,\ldots,n_K)$, upon observing the action a_j taken by the opponent at time step $t+1$, agent i can update the belief as $f_{t+1}(\lambda;Q+Q_{a_{j,k}},n_0,n_1,\ldots,n_k+1,\ldots,n_K)$.

Proof:

$$
\begin{aligned}
f(\lambda|a_{j,k}) &= \frac{P(a_{j,k}|\lambda)f(\lambda;Q,n_0,n_1,\ldots,n_K)}{\int_0^\infty P(a_{j,k}|\lambda')f(\lambda';Q,n_0,n_1,\ldots,n_K)d\lambda'} \\
&= \frac{\frac{e^{\lambda Q_{a_{j,k}}}}{\sum_{l=1}^m e^{\lambda Q_{a_{j,k}}}} \cdot \frac{e^{\lambda Q}\cdot g(Q,n)}{\prod_{k=0}^K (\sum_{l=1}^m e^{\lambda Q_{a_{l,k}}})^{n_k}}}{\int_0^\infty \frac{e^{\lambda' Q_{a_{j,k}}}}{\sum_{l=1}^m e^{\lambda' Q_{a_{j,k}}}} \cdot \frac{e^{\lambda' Q}\cdot g(Q,n)}{\prod_{k=0}^K (\sum_{l=1}^m e^{\lambda' Q_{a_{l,k}}})^{n_k}} d\lambda'} \\
&= \frac{\frac{e^{\lambda(Q_{a_{j,k}}+Q)}\cdot g(Q,n)}{\prod\limits_{\substack{i=0\\ i\neq k}}^K (\sum_{l=1}^m e^{\lambda Q_{a_{l,k}}})^{n_k}\cdot(\sum_{l=1}^m e^{\lambda Q_{a_{l,k}}})^{n_k+1}}}{\int_0^\infty \frac{e^{\lambda'(Q_{a_{j,k}}+Q)}\cdot g(Q,n)}{\prod\limits_{\substack{i=0\\ i\neq k}}^K (\sum_{l=1}^m e^{\lambda' Q_{a_{l,k}}})^{n_k}\cdot(\sum_{l=1}^m e^{\lambda' Q_{a_{l,k}}})^{n_k+1}} d\lambda'} \\
&= f(\lambda;Q+Q_{a_{j,k}},n_0,n_1,\ldots,n_k+1,\ldots,n_K)
\end{aligned}
$$

where we denote $\frac{1}{\int_0^\infty e^{\lambda' Q}/\prod_{k=0}^K (\sum_{l=1}^m e^{\lambda' Q_{a_{l,k}}})^{n_k} d\lambda'}$ as $g(Q,n)$ for simplicity.

When the distribution of the continuous variable λ at the next time step s_{t+1} is obtained, the expectation of rationality degree can be calculated using the following equation:

$$
\mathbb{E}(\lambda) = \int_0^\infty \lambda f(\lambda;Q+Q_{a_{j,k}},n_0,n_1,\ldots,n_k+1,\ldots,n_K)d\lambda \tag{4.8}
$$

The above describes how to update the knowledge of the opponent's reasoning level k and rationality degree λ at each time step. Algorithm 5 provides a clearer understanding of the process.

Through dynamic belief updates, AVs can more accurately predict pedestrian behavior and adjust their strategies based on the latest beliefs. For example, the AV initially regards its opponent as a level-1 pedestrian and maintains its speed accordingly based on this expectation. However, when it observes the pedestrian proceeding to cross the road, the AV increases the probability of the pedestrian being level-0 while reducing the rationality value. This dynamic update mechanism enables AVs to better adapt to changing environments, thereby behaving more intelligently and human-like.

Algorithm 3: Belief Update

Input: s_t: Current state, a_j: Observed opponent's action, $Q_{a_{j,k}}$: Opponent's action a_j's expected utility for each level k, $b_{k,t}$: Prior belief about reasoning level, $P(s_{t+1}, a_j|k)$: Probability for a_j at each level k, $f(\lambda; Q, n_0, n_1, \ldots, n_K)$: Prior distribution about rationality

Output: Updated belief $b_{k,t+1}$, $b_{\lambda,t+1}$

for $k = 0$ **to** $K-1$ **do**

$P(k|s_{t+1}, a_{j,t+1}) = \frac{P(s_{t+1}, a_{j,t+1}|k) b_{k,t}(k)}{\sum_{k' \in \Theta} P(s_{t+1}, a_{j,t+1}|k') b_{k,t}(k')}$

$b_{k,t+1}(k) \leftarrow P(k|s_{t+1}, a_{j,t+1})$;

$k \leftarrow \arg\max(b_{k,t+1})$;

$\mathbb{E}(\lambda) \leftarrow \int_0^\infty \lambda f(\lambda; Q + Q_{a_{j,k}}, n_0, \ldots, n_k + 1, \ldots, n_K) d\lambda$;

$b_{\lambda,t+1} \leftarrow \mathbb{E}(\lambda)$;

return $b_{k,t+1}$, $b_{\lambda,t+1}$;

4.3.3 Monte Carlo Tree Search

Monte Carlo Tree Search is a heuristic search algorithm widely used for sequential decision-making problems, particularly in environments characterized by uncertainty and complex interactions. It predicts potential future outcomes by running numerous simulations and guides the selection of optimal actions based on accumulated rewards. The algorithm proceeds through four main phases: selection, expansion, simulation, and backpropagation [227]. In our study, MCTS is employed as the core mechanism to evaluate the expected utility of each candidate action for both the AV and pedestrian agents at each cognitive reasoning level for each time step (Fig. 4.1d, Fig. 4.1h). MCTS is selected because it provides forward-looking planning under uncertainty, enabling the agent to simulate and evaluate multiple possible future interaction trajectories arising from its opponent's stochastic behavior. By simulating possible future interaction trajectories and propagating the resulting outcomes through a decision tree, MCTS allows each agent to iteratively refine its strategy and identify the action that offers the highest expected return.

At each decision step, the MCTS algorithm is initialized from the current environment state s_t. During the selection phase, the algorithm traverses the existing search tree by selecting child nodes that maximize the Upper Confidence Bound for Trees (UCT) score, which balances exploitation of high-reward nodes and exploration of less-visited ones. The UCT score for each node is calculated using the following formula [228]:

$$\mathrm{UCT}(s,a) = \bar{Q}_a + C\sqrt{\frac{\ln N_s}{N_a}} \tag{4.9}$$

where $\bar{Q}_a$ is the average utility of action a at state s_t, N_s is the total number of visits to the state s, N_a is the number of times action a was chosen, and C is a constant that balances exploration and exploitation.

In the expansion phase, new decision nodes are generated for unexplored actions available to the AV and pedestrian, reflecting different trajectories the agents might follow from the current state. The simulation step starts from the newly expanded node, where the actions of the agent and its opponent are randomly simulated in turn until reaching the terminal state or the maximum search depth. In this stage, a random strategy is used to select actions and simulate the opponent's strategies, estimating the potential value of the node. Notably, when calculating the optimal strategy for the agent at level-k, its opponent follows the policy of level-$k-1$; for level-0 agents, their opponents are regarded as static obstacles. Upon simulation completion, the observed reward is propagated back up the tree during the backpropagation phase. Each visited node updates its visit count and cumulative reward statistics accordingly. Repeating this process over many iterations leads to a well-populated decision tree, where each node stores reliable estimates of future utility.

Through this process, MCTS evaluates the potential effects of different actions through extensive simulations without relying on specific domain knowledge, enabling AV to make efficient and safe decisions in complex and uncertain environments.

4.3.4 Reward Function Design

In the expansion stage of MCTS, each newly expanded node represents a specific state-action pair, and its value is estimated by simulating the resulting trajectory and accumulating the rewards obtained along the way. This reward plays a central role in evaluating the utility of each simulated path and is therefore crucial for guiding the search process.

In our framework, the reward function is explicitly designed to reflect the AV's decision-making priorities in pedestrian-involved environments. It encourages goal-directed behavior while penalizing unsafe, inefficient, or uncomfortable actions. The detailed reward formulation is as follows:

$$r = w_1 \cdot r_{\text{goal}} + w_2 \cdot r_{\text{collision}} + w_3 \cdot r_{\text{efficiency}} + w_4 \cdot r_{\text{comfort}} \tag{4.10}$$

where, w_1, w_2, w_3, w_4 are weights that determine the relative importance of each reward component. The detailed explanation of each reward is presented below.

Goal reward r_{goal}: A positive reward of 10 is given when the AV reaches the goal region:

$$r_{\text{goal}} = \begin{cases} 10, & \text{if goal reached} \\ 0, & \text{otherwise} \end{cases} \tag{4.11}$$

Collision penalty $r_{\text{collision}}$: If a collision occurs, the AV receives a strong penalty of -20, with an additional penalty proportional to its current speed, thereby discouraging high-speed collisions.

$$r_{\text{collision}} = \begin{cases} -20 - 1 \cdot v_t, & \text{if collision with pedestrians} \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

Efficiency penalty $r_{\text{efficiency}}$: This term encourages the AV to maintain a reasonable and efficient speed while avoiding unnecessary stops. The reward is defined as:

$$r_{\text{efficiency}} = \begin{cases} -5, & \text{if v} = 0 \\ -d/v, & \text{otherwise} \end{cases} \tag{4.13}$$

where v denotes the AV's current speed, and d represents the distance to the goal.

Comfort penalty r_{comfort}: To encourage smooth driving, the large changes in acceleration between time steps is introduced:

$$r_{\text{comfort}} = -1 \cdot |a_t - a_{t-1}| \tag{4.14}$$

By jointly considering safety, efficiency, and comfort in the reward formulation, the AV is guided to make well-balanced decisions that prioritize collision avoidance, maintain smooth and efficient motion, and ensure passenger comfort.

4.4 Experiments and Results

4.4.1 Experiment Setup

We conducted a series of simulation experiments for verification to evaluate the effectiveness of the AV model and decision-making algorithm we developed. We built a simulation scenario where the AV interacts with a pedestrian crossing the road. The AV is 5 m long and 2 m wide, driving on a single-lane road that is 3.65 m wide. The pedestrian's goal is to cross the road to reach the opposite side.

We previously conducted experiments on real human drivers and pedestrians interacting in a VR environment. During these experiments, the driver continuously drove along the road at a random initial speed and interacted with the pedestrian crossing the road. Since VR technology providing reliable and essential data has been widely applied in the interactive research and testing of AVs [24], [229], we are expected to validate and evaluate our model by comparing our simulation data with the VR experimental data. This experimental data includes the position and speed for both the vehicle and the pedestrian, as well as acceleration for the vehicle only. We randomly selected 100 scenarios from these experimental data for this simulation experiment. The initial conditions of each scenario were input into our model for simulation, and each scenario was simulated 100 times to ensure the reliability and statistical significance of the results.

The initial data included the initial positions of the vehicle and the pedestrian, the vehicle's initial velocity and acceleration, and the pedestrian's velocity. In our models, we assume that the AV and pedestrian follow a straight line with only longitudinal movements. The acceleration range of the AV is set to [-5, 5] m/s^2, and the speed range of the pedestrian is [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2] m/s. We assume that the maximum reasoning ability of the agent is $k\_\max = 2$, with an initial rationality degree value of $E(\lambda) = 10$. The level-0 policy is non-strategic, which believes that the agent calculates its optimal strategy based on the assumption that its opponents are static obstacles. Our experiments were conducted on an Intel i7 processor without GPU acceleration. Through multiple tests, we observed that the average decision-making time was slightly below 0.8 seconds, and therefore selected 0.8 s as the time step. While the empirical runtime on our CPU-based setup results in a 0.8 s decision interval, the computational

load is dominated by the MCTS component, whose cost grows with rollout numbers and search depth. The belief-update and quantal-response computations remain lightweight. In practice, the decision cycle can be accelerated through optimization strategies such as reducing rollout counts or tree depth, or by employing faster hardware.

4.4.2 Results

This section will verify and evaluate our proposed model's performance through qualitative, quantitative analysis, as well as sensitive analysis of the initial bounded rationality value.

4.4.2.1 Qualitative Analysis

Three simulation examples will be given to clearly and intuitively demonstrate how our AV model operates under different scenarios.

Case 1: This case illustrates an AV yielding to a pedestrian, randomly selected from a set of 100 scenarios. Initially, the AV is 46.309 m away from the pedestrian, traveling at a speed of 9.348 m/s with an acceleration of -0.11 m/s^2 towards the pedestrian. The pedestrian is located 2.564 m laterally from the AV on the road. Fig. 4.2 shows the interaction process, combining visual snapshots of their positions with detailed speed data. Information about the state evolution of the pedestrian and vehicle is provided in Fig. 4.3.

Under these initial conditions, the AV calculates a higher probability that the pedestrian is a level-0 agent, indicating that the pedestrian is more likely to cross the road. Therefore, the AV decides to decelerate to avoid collision. When the AV observes the pedestrian stepping into the lane at a speed of 1.2 m/s, it executes a more pronounced deceleration. At the same time, the AV updates its assessment of the pedestrian's rationality based on their actions, concluding that the pedestrian remains in a rational state.

As the pedestrian continues to cross and approach its destination, the AV gradually reduces its deceleration and eventually transitions back to acceleration. Notably, the AV does not come to a complete stop but maintains a reduced speed while the pedestrian crosses. This entire process demonstrates the AV's ability to yield to pedestrians by making real-time decisions, ensuring both safety and efficiency.

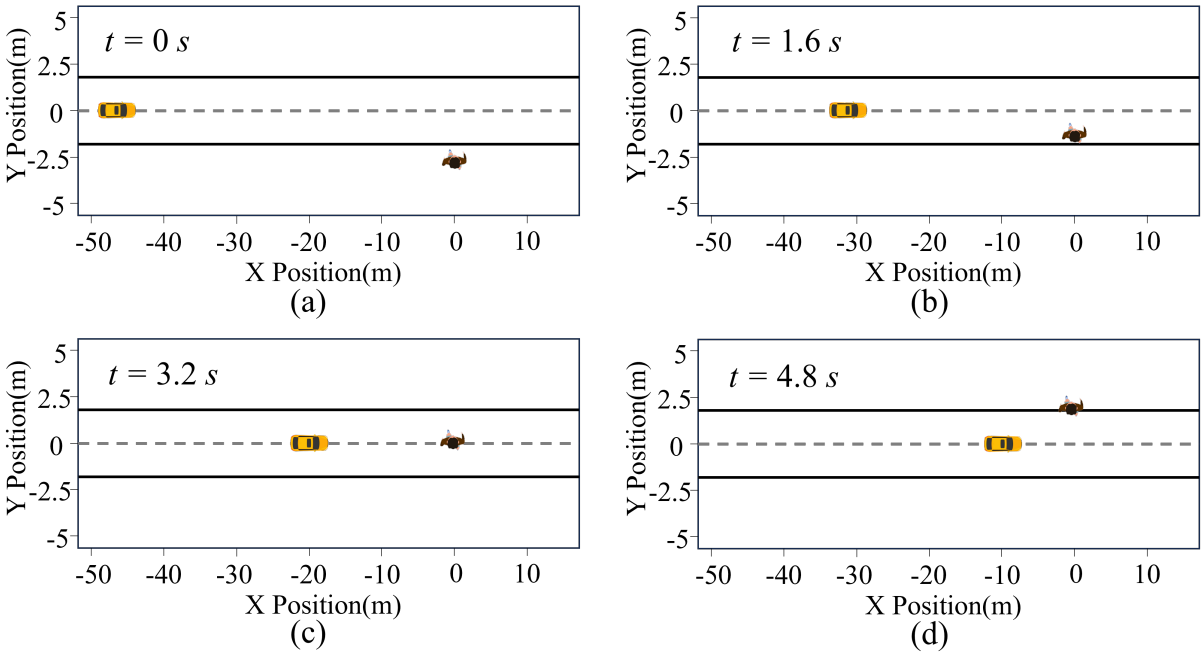


Figure 4.2: Simulation of interaction process in case 1. (a) State: t = 0 s, v_{ped} = 0.03 m/s, v_{AV} = 9.348 m/s, a_{AV} = -0.11 m/s^2. (b) State: t = 1.6 s, v_{ped} = 1.2 m/s, v_{AV} = 8.31 m/s, a_{AV} = -1.397 m/s^2. (c) State: t = 3.2 s, v_{ped} = 1.2 m/s, v_{AV} = 6.72 m/s, a_{AV} = -0.111 m/s^2. (d) State: t = 4.8 s, v_{ped} = 0.7 m/s, v_{AV} = 6.57 m/s, a_{AV} = -0.283 m/s^2.

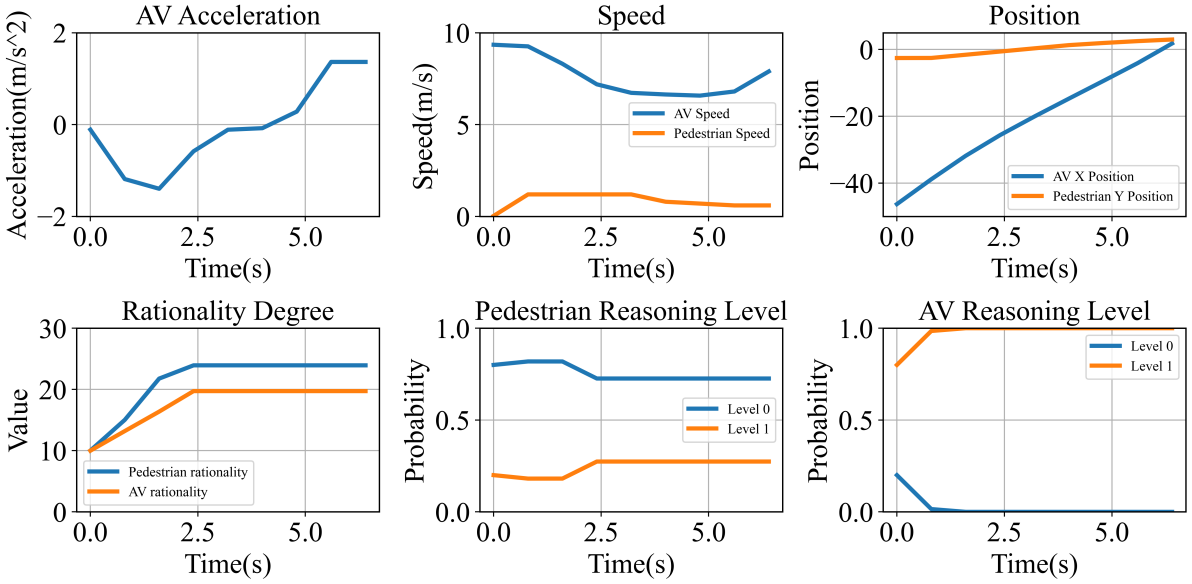


Figure 4.3: State evolution in the simulation of case 1.

Case 2: This case illustrates a scenario where a pedestrian yields to an AV. Compared to case 1, the initial longitudinal distance between the AV and the pedestrian is updated to 34 m with other conditions remaining the same. The interaction process and detailed evolution states are shown in Fig. 4.4 and Fig. 4.5.

At the first time step, the AV is 34 m away, moving at a speed of 9.348 m/s, and an acceleration of -0.11 m/s^2. The pedestrian is positioned on the road at a lateral distance of 2.564 m from the AV. Unlike case 1, the AV determines that the pedestrian poses a lower probability of crossing and designates the pedestrian as a level-1 agent, signifying that the pedestrian will not recklessly enter the lane and yield to AV.

As a result, the AV opts to maintain a slight deceleration in case the pedestrian suddenly changes his mind and steps into the lane. However, after observing that the pedestrian does not appear in the lane, the AV confirms that the pedestrian will indeed not cross and perceives its behavior as rational. At this point, the AV sustains a slight deceleration to ensure the pedestrian's safety and maintain smooth operation without significant speed adjustments.

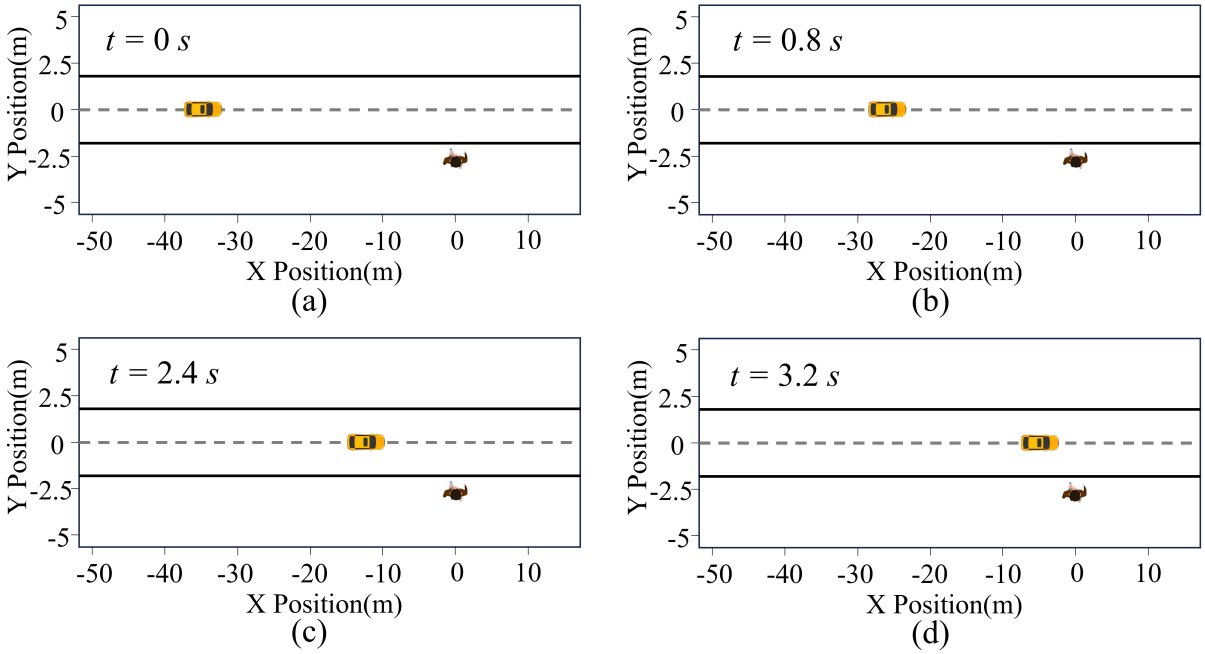


Figure 4.4: Simulation of interaction process in case 2. (a) State: t = 0 s, v_{ped} = 0.03 m/s, v_{AV} = 9.348 m/s, a_{AV} = -0.11 m/s^2. (b) State: t = 0.8 s, v_{ped} = 0 m/s, v_{AV} = 9.26 m/s, a_{AV} = -0.33 m/s^2. (c) State: t = 2.4 s, v_{ped} = 0 m/s, v_{AV} = 8.83 m/s, a_{AV} = -0.216 m/s^2. (d) State: t = 3.2 s, v_{ped} = 0.1 m/s, v_{AV} = 8.62 m/s, a_{AV} = -0.175 m/s^2.

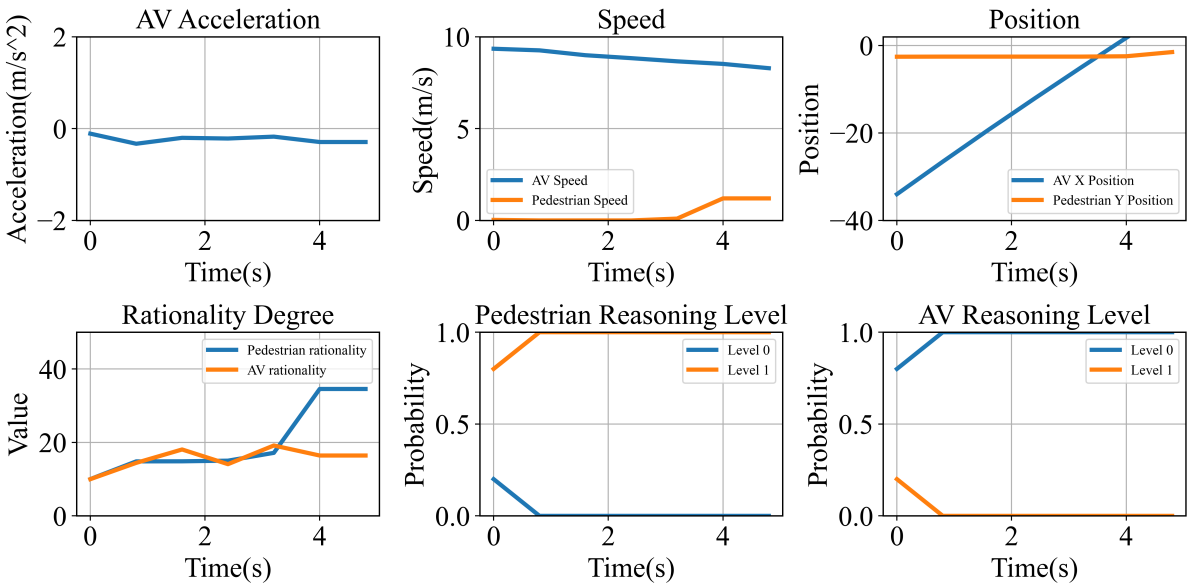


Figure 4.5: State evolution in the simulation of case 2.

Case 3: This case primarily aims to validate the AV's ability to identify irrational pedestrian behavior. In the initial conditions of Case 2, the AV deemed the pedestrian's decision to yield as rational, and the pedestrian model indeed waited for the AV to pass before crossing the road. In this case, we maintain the same initial conditions as in Case 2, with the only difference being the replacement of the pedestrian model with custom-design action for the pedestrian. Specifically, we program the pedestrian to cross the road under these conditions at a speed of 1.2 m/s.

As depicted in Fig. 4.7, the AV, similar to Case 2, initially assumes a low likelihood of pedestrians crossing the road and thus only slows down slightly. However, when the pedestrian begins moving, the AV updates its estimation of the pedestrian being a level-0 agent based on their actions. At the same time, a marked drop in the pedestrian's rationality value indicates that the AV deems it irrational for the pedestrian to cross the road under the current circumstance. According to these judgments, the AV decelerates more sharply than in Case 2, resulting in a rapid decrease in vehicle speed. This indicates that the proposed AV model can effectively update its understanding of pedestrian rationality based on real-time behaviors and appropriately adjust its acceleration to prevent potential collisions.

In conclusion, the AV's ability to dynamically adjust its perception of pedestrian behavior and response with suitable deceleration shows a good performance of the AV model in various pedestrian interaction scenarios.

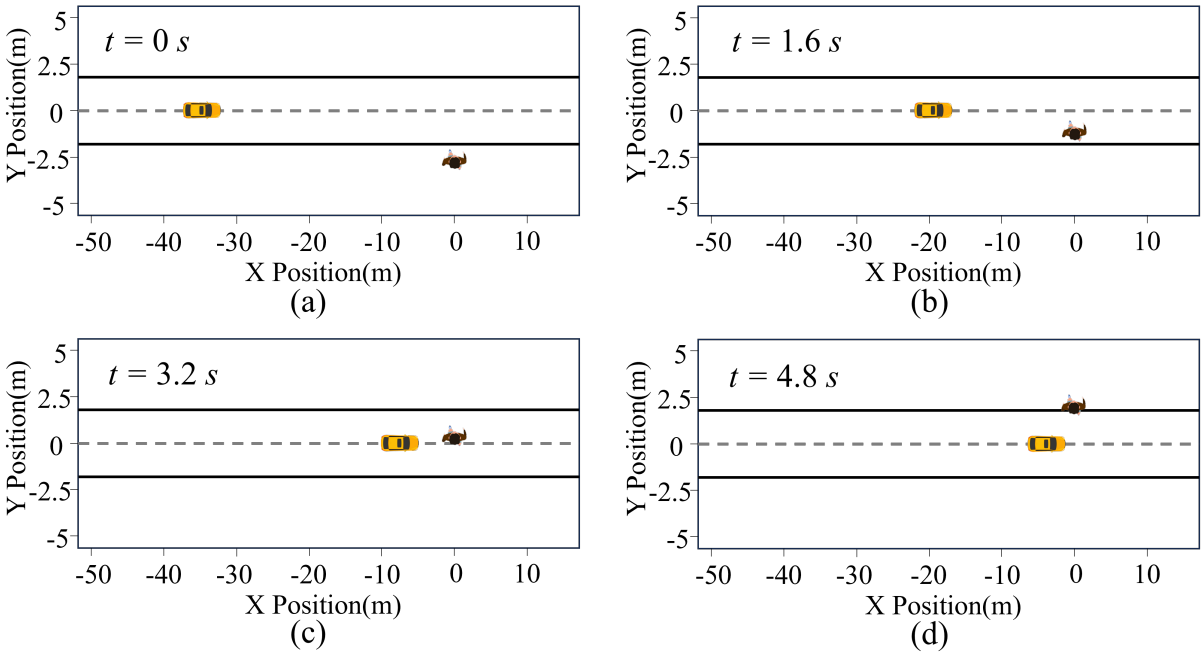


Figure 4.6: Simulation of interaction process in case 3. (a) State: t = 0 s, v_{ped} = 0.03 m/s, v_{AV} = 9.348 m/s, a_{AV} = -0.11 m/s^2. (b) State: t = 1.6 s, v_{ped} = 1.2 m/s, v_{AV} = 8.99 m/s, a_{AV} = -3.081 m/s^2. (c) State: t = 3.2 s, v_{ped} = 1.2 m/s, v_{AV} = 4.83 m/s, a_{AV} = -1.646 m/s^2. (d) State: t = 4.8 s, v_{ped} = 1.2 m/s, v_{AV} = 2.05 m/s, a_{AV} = -1.641 m/s^2.

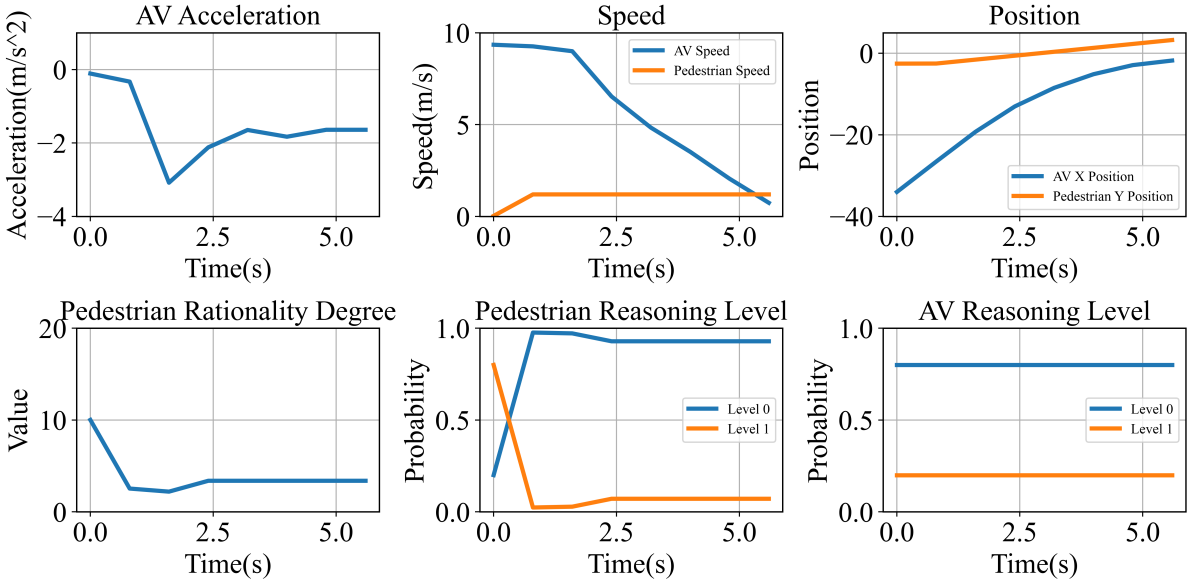


Figure 4.7: State evolution in the simulation of case 3.

4.4.2.2 Quantitative Analysis

Following the analysis method in the work [230], we will conduct the quantitative evaluation from three aspects: safety, efficiency, and smoothness. Due to the mismatch in required input variables between our framework and existing datasets, baseline models cannot be meaningfully applied. Therefore, real human behavior from the VR experiment is used as the primary reference for evaluating the proposed approach. We input 100 scenarios into our model for simulation, with each scenario being simulated 100 times. Results in Table 4.2 show the comparison between our proposed method with VR experiment driving data. For safety, our driver model has a slightly higher collision rate compared to the VR experiment data. The collision rate is not zero because the proposed framework introduces stochasticity both in the agents' decision-making process and in the planning procedure. Probabilistic action selection in the quantal response model, together with stochastic sampling in the Monte Carlo simulation, may occasionally produce rare combinations of actions that result in collisions. Therefore, the observed 0.15% collision rate reflects the system's behavior under uncertainty.

Table 4.2: Statistic results of our proposed approach compared with VR experiments.

	Method	VR experiment	Ours
Safety	Collision rate	/	**0.15%**
Efficiency	Yielding rate	51%	**75.03%**
	Average vehicle speed (m/s)	9.468	**9.600**
Smoothness	Average vehicle jerk	0.897	**0.843**
	Average maximum absolute acceleration/deceleration (m/s^2)	2.228	**1.867**

To evaluate driving efficiency, the vehicle yielding rate and average vehicle speed are considered. The yielding rate observed in the simulation (75.03%) is higher than in the VR experiment (51%). This suggests that the algorithm is more cautious, which may slightly affect the average speed. However, the average speed in the simulation (9.600 m/s) is slightly faster compared to the experiment (9.468 m/s), indicating that our approach can maintain efficiency even with a greater yielding rate. Fig. 4.8a compares the average vehicle speed across 100 different scenarios. The VR experimental driving data and the simulation data follow similar trends, with peaks and valleys appearing in roughly the same places. Although specific speed values fluctuate in individual scenarios, the average speeds in both the VR experiment and the simulation remain around similar values.

Two parameters, jerk and maximum absolute acceleration/deceleration, are used to evaluate smoothness. The lower average jerk value and average maximum absolute acceleration/deceleration value in Table 4.2 indicate that our proposed method can achieve smoother and more comfortable driving behavior. Fig. 4.8b and Fig. 4.8c compare the

average jerk and maximum absolute acceleration/deceleration of VR experimental data and simulation data across different scenarios. In general, the simulation data shows less fluctuation and lower values, showing more stable acceleration/deceleration control. The overall trends and average values suggest that our approach performs better than the experimental data in driving smoothness.

In summary, the quantitative analysis shows that the proposed AV decision-making algorithm performs well in safety, efficiency, and smoothness. The similar average vehicle speed values between VR experimental data and simulation data indicate that our algorithm successfully captures key aspects of the driving behavior exhibited by human drivers. Additionally, the lower jerk values and maximum absolute acceleration/deceleration values in our simulations suggest that our method achieves smoother driving compared to the experimental data.

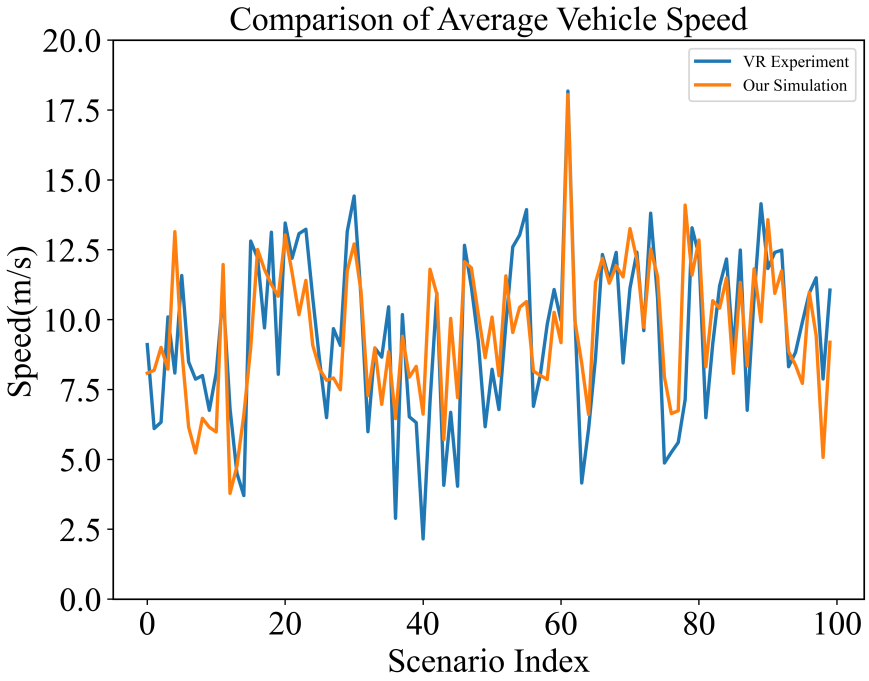
4.4.2.3 Sensitive Analysis

We conducted a sensitive analysis on the model's initial bounded rationality value, a key parameter in our dynamic updating mechanism for rationality. It is noted that the bounded rationality is characterized by the parameter λ, which is not a fixed value but follows a probability distribution that reflects the uncertainty about the opponent's rationality degree. In our analysis, the value of the bounded rationality therefore specifically refers to the expectation value of the lambda based on this distribution. Using the default initial value of $E(\lambda)$ as 10, we also trained and evaluated the model with values that were 1/10 and 10 times the default (see Table 4.3).

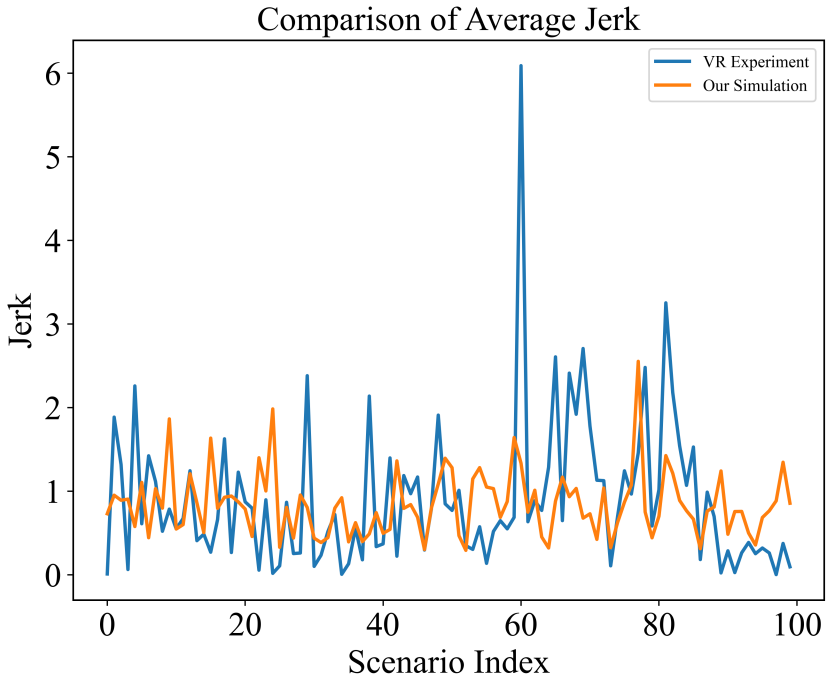
Table 4.3: Sensitive analysis of initial bounded rationality.

Initial bounded rationality	Collision rate	Yielding rate	Average vehicle speed (m/s)	Average vehicle jerk	Average maximum absolute acceleration/deceleration (m/s^2)
1	2.76%	72.52%	9.259	0.928	2.022
10	**0.15%**	**75.03%**	**9.600**	**0.843**	**1.867**
100	0.08%	74.57%	9.935	0.748	1.717

When the AV estimates the pedestrian's initial rationality value as low ($E(\lambda)$ =1), it shows the lowest average vehicle speed, the highest jerk, and the largest maximum absolute acceleration/deceleration, along with a significantly higher collision rate. This occurs because the AV perceives the pedestrian's behavior as more random and unpredictable under this assumption, prompting it to adopt a conservative strategy, such as reducing or maintaining its speed, to minimize collision risk. However, this unpredictability requires more frequent and significant deceleration adjustments to handle potential sudden changes, resulting in increased jerk and acceleration/deceleration values. Additionally, the difficulty in accurately predicting pedestrian behavior further increases the likelihood of misjudgment, contributing to the elevated collision rate.

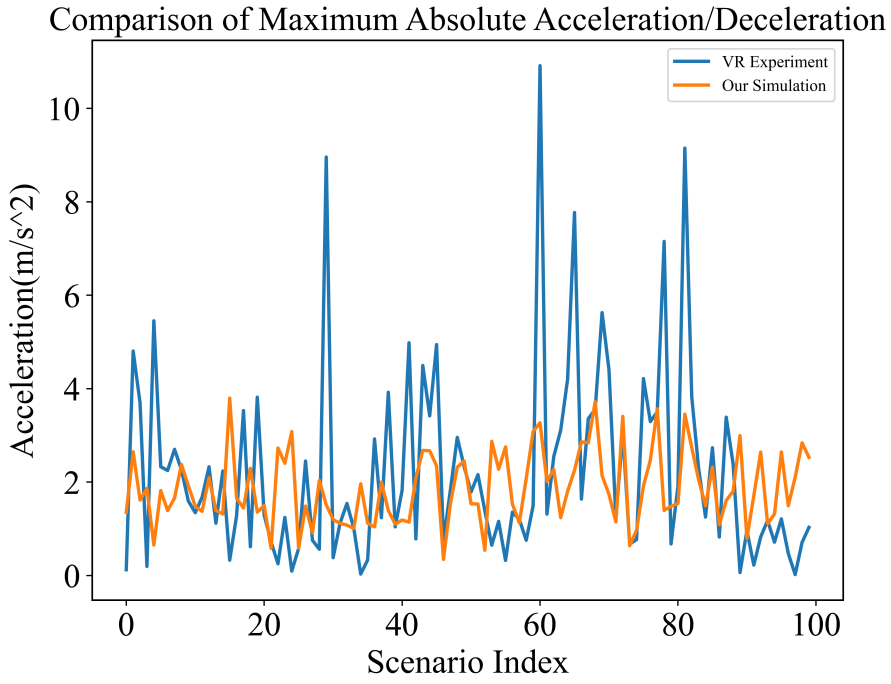


(a)



(b)

Figure 4.8: Comparison across different scenarios from average vehicle speed, average vehicle jerk and maximum absolute vehicle acceleration/deceleration between the VR experiment and simulation data.



(c)

Figure 4.8: (Continued) Comparison across different scenarios from average vehicle speed, average vehicle jerk and maximum absolute vehicle acceleration/deceleration between the VR experiment and simulation data.

In contrast, the highest initial rationality value in Table 4.3 results in a safer, more efficient and smoother motion. This belief enables the AV to better predict pedestrian behavior and execute its planning strategy with greater confidence. Consequently, the AV operates fewer frequent or abrupt trajectory adjustments, leading to smoother acceleration and deceleration patterns that significantly reduce jerk and acceleration values. Moreover, the enhanced ability to predict pedestrian behavior allows the AV to act more decisively, optimizing navigation efficiency while further minimizing collision risks.

Overall, the set of the initial rationality value can influence the AV's strategy selection and adjustment capabilities. While this value is dynamically updated during continuous interactions, such adjustments require time and multiple observations to become significant. In shorter interaction scenarios, the impact of the initial value may persist throughout the process, with limited room for dynamic updates.

Although assuming an initial rationality value of 100 achieves the best performance, an excessively high value can make the AV overly confident in predicting pedestrian behavior, thereby limiting its dynamic adjustment capability in real environments. In reality, pedestrian behavior in practice often involves uncertainty. A lower initial value (10) provides the AV with greater flexibility for dynamic adjustments, allowing it to better adapt to such uncertainties. Therefore, we ultimately chose 10 as the initial rationality value to balance performance and adaptability.

4.5 Chapter Conclusion

This chapter first identified the limitations in current studies on AV–pedestrian interactions. Many previous studies rely on statistical methods or treat interactions as one-time events, which fail to capture the dynamic and unpredictable nature of pedestrian behaviors and are unable to represent behavioral uncertainty effectively. Meanwhile, game-theoretic approaches, often assume that all agents are fully rational and always act according to Nash equilibrium. Such idealized assumptions diverge from real-world conditions, where human decision-making is influenced by cognitive limitations, resulting in inaccurate behavior modeling. To address these gaps, this study proposes an innovative framework to address the decision-making challenges AVs face when interacting with pedestrians at the unsignalized intersection. First, we integrate the POMDP with behavioral game theory to model these interactions, capturing the uncertainty and dynamism of pedestrian behavior. Second, both the AV and pedestrian are modeled as DB-QCH models, accounting for human reasoning limitations and bounded rationality, thus enabling more realistic interaction simulations compared to traditional game theory approaches. Moreover, the dynamic updating mechanism for the opponent's rationality degree is introduced, which allows the AV to adjust its strategies based on real-time observations. Finally, a trained neural network is developed to guide MCTS within the AV's continuous action space, improving decision-making efficiency and effectiveness.

Simulation results demonstrate that our method excels in safety, efficiency, and smoothness, closely capturing the key features of the driving behavior operated by real human drivers in VR experiments. One limitation of this work is that the reward function incorporates scenario-specific weight parameters and may require re-calibration when deployed in new environments. Future work should investigate more generalizable or automatically optimized reward designs to enhance applicability across diverse scenarios. Moreover, this work does not conduct a sensitivity analysis of the weight parameters in the reward function. Future work should include a systematic sensitivity analysis of the weights to further assess model generalizability and robustness. Although our model performs well,

the current research is limited to a simple scenario of a single AV and a single pedestrian interaction. In the next chapter, we will expand our scope to include the interaction between a single AV and multiple pedestrians, allowing the proposed AV decision-making algorithm to handle more complex scenarios. Our work will explore the AV's decision-making in multi-pedestrian interactions, addressing challenges such as increased computational demands and behavioral uncertainty through multi-agent methods and uncertainty modeling.

Chapter 5

Free Energy-Inspired Cognitive Risk Integration for AV Navigation in Pedestrian-Rich Environments

5.1 Background

This chapter focuses on the decision-making problem of AVs interacting with multiple pedestrians in shared spaces, as shown in Fig. 1.1b, aiming to develop a safe and efficient decision-making model capable of handling multi-agent interactions and complex environmental uncertainties.

While numerous studies have explored crowd navigation for small-scale mobile robots [231], [232], relatively few have focused on AVs operating in similar pedestrian-rich environments. Although AVs represent a specialized class of robotic systems, their navigation differs from that of smaller robots. Unlike service robots, AVs have larger physical footprints, operate at higher speeds, and are subject to more complex kinematic and safety constraints [233]. For example, a collision involving a service robot might cause minor disruption, whereas an AV-pedestrian collision could have life-threatening consequences. These distinctions make it impractical to directly apply robot navigation strategies to AVs. As urban environments increasingly require AVs to coexist with pedestrian crowds, there is a growing need for decision-making frameworks that support safe and efficient navigation in such complex settings.

Accurate pedestrian modeling plays a critical role in enabling realistic and adaptive AV decision-making in interactive environments. Many studies have adopted overly simplified pedestrian models with fixed behavior patterns, such as constant velocity assumptions [12] or pre-recorded real-world trajectories [13], [234]. These approaches commonly treat

pedestrians as passive agents who are unable to dynamically respond to the AV's behavior during simulation. Consequently, they fail to capture the bidirectional adjustment mechanisms essential for realistic interaction modeling. Consider, for example, a scenario where an AV and a pedestrian approach a potential conflict point simultaneously. The AV might yield, expecting the pedestrian to proceed. However, if the pedestrian follows a predefined path without perceiving or reacting to the AV's action, they may also hesitate or stop, leading to unnatural and unrealistic interaction outcomes. This limitation weakens the ability to model dynamic adaptive behaviors, which are essential for capturing the nature of human-vehicle interactions in shared environments.

To address these limitations, our work proposes a graph-enhanced deep reinforcement learning framework for modeling interactions between an AV and multiple pedestrians. Fig. 5.1 illustrates the overall architecture of our proposed framework in urban shared spaces. In this work, we adopt the SAC algorithm to train the AV's decision-making policy. Specifically, a GCN module is integrated into both the actor and critic networks to capture structured multi-agent interactions. The GCN operates on a dynamically constructed interaction graph, whose adjacency matrix encodes the interaction intensity between agents. The edge weights are computed using a combined risk metric that integrates physical risk and cognitive uncertainty, with the latter derived from the Free Energy Principle. Furthermore, pedestrians are modeled as intelligent agents with both interactive capabilities and cognitive awareness. Their behavior is generated through the proposed CR-SFM, which employs the same integrated risk signal used in the GCN to modulate the contribution of the goal-directed and repulsive forces. By incorporating both physical and cognitive risk into the force computation, pedestrian trajectories can be adaptively adjusted in response to surrounding dynamics, yielding more realistic and human-like behaviors in complex environments.

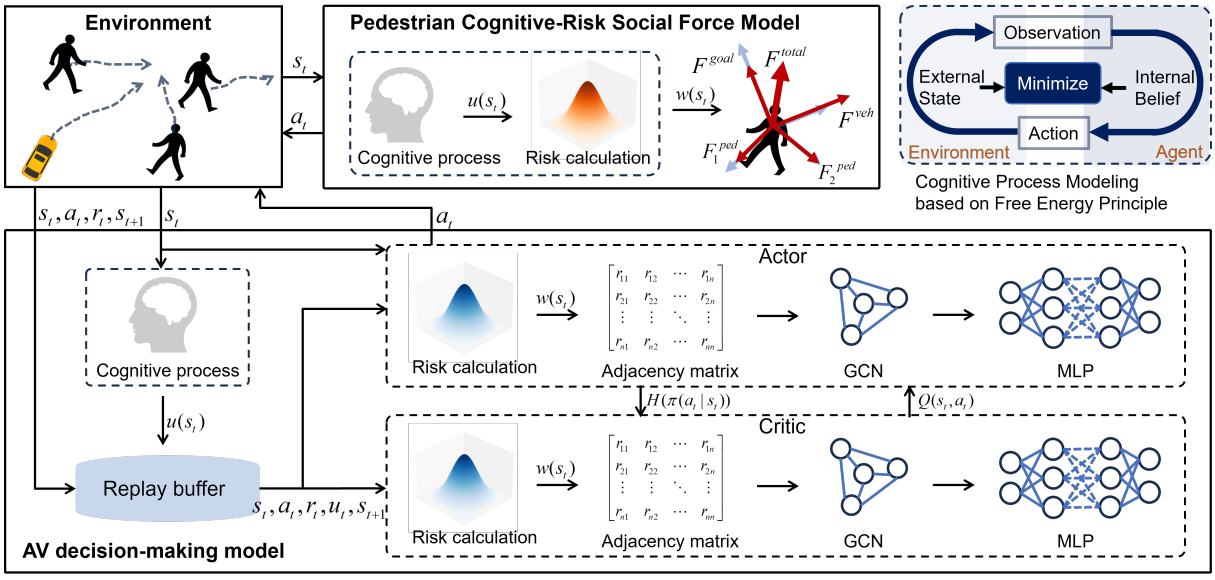


Figure 5.1: The proposed framework of AV-pedestrian interaction in the urban shared space.

In order to better understand the interactions and resolve navigation challenges in crowded environments, the main contributions of this work are as follows: 1). A cognitive uncertainty modeling approach based on the free energy principle is introduced to characterize subjective human prediction biases and the iterative adjustment process when reasoning about the behaviors of other agents. 2). A pedestrian model called the Cognitive-Risk Social Force Model is proposed, which dynamically adjusts the weights of goal-directed and repulsive forces based on a fused risk measure. This enables pedestrian models to autonomously adapt their motions in uncertain and interactive scenarios, achieving more realistic behavior than conventional models with fixed coefficients. 3). A risk-encoded interaction graph is constructed to improve AV decision-making within a reinforcement learning framework. The combined risk is used to define the adjacency matrix of a GCN, which is embedded in both the actor and critic networks of the SAC framework. This integration allows the policy to better perceive interactive relationships and potential threats, thus improving decision-making performance.

5.2 Problem Formulation

This work focuses on the AV navigation problem in environments with multiple pedestrians. The objective is to develop a continuous decision-making strategy that ensures safe and efficient operation amid dynamic human interactions.

We model this problem as a MDP, defined by the tuple: $\langle \mathscr{S}, \mathscr{A}, \mathscr{P}, \mathscr{R}, \gamma \rangle$. The state space $\mathscr{S}$ captures the dynamic state of the environment at each time step t, denoted by $s_t \in \mathscr{S}$. We assign index 0 to the AV and indices 1 to n to the surrounding pedestrians. The AV's state s_t^0 includes its current position (x, y), heading angle θ, velocity v, acceleration α, goal position $(x_{\text{goal}}, y_{\text{goal}})$, distance to the goal position d_{goal}, and angular deviation from the goal direction $\Delta\theta_{\text{goal}}$. Each pedestrian's state s_t^i, for $i = 1, \ldots, n$, consists of their position, heading angle, and walking velocity. The whole environment state at time t is defined as $s_t : (s_t^0, s_t^1, \ldots, s_t^n)$. The action space $\mathscr{A}$ represents the set of possible actions available to the AV. Each action $a_t \in \mathscr{A}$, is defined as a tuple $a_t = [\alpha, \Delta\theta]$, where α denotes the acceleration and $\Delta\theta$ denotes the change in heading direction. The state transition probability $\mathscr{P}(s_{t+1}|s_t, a_t)$ defines how the environment evolves from the current state s_t to next state s_{t+1} given the action taken by the AV. Following the transition, the AV receives a reward $r_t \in \mathscr{R}$, indicating the immediate utility of its chosen action.

The AV model aims to learn a policy $\pi(a_t|s_t)$ that produces a sequence of optimal actions to maximize the expected cumulative discounted reward over time, where the process terminates upon reaching the destination. This decision-making problem can thus be defined as follows:

$$\begin{aligned} \max_{\pi} \quad & \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t) \mid a_t \sim \pi\right] \\ \text{subject to} \quad & s_{t+1} \sim \mathscr{P}(s_{t+1} \mid s_t, a_t), \\ & a_t \in \mathscr{A}, s_t \in \mathscr{S}, \\ & r_t \in \mathscr{R}, \forall t \in \mathbb{N} \end{aligned} \tag{5.1}$$

where $\gamma \in [0,1]$ is the discount factor that determines the relative importance of future rewards.

5.3 Methodology

This section describes the methodology for modeling AV-pedestrian interactions in crowded environments and addressing the previously formulated problem.

5.3.1 Vehicle Dynamics

In this work, the AV is modeled using a point-mass kinematic model to represent its motion dynamics. The control inputs to this model are the acceleration a_t and the change in heading angle $\Delta\theta_t$, both of which are directly produced by the AV's decision-making system. The following equations define the vehicle's state transition over time:

$$\begin{aligned} x_{t+1} &= x_t + v_t \cdot \cos(\theta_t) \cdot \Delta t + \frac{1}{2} \cdot a_t \cdot (\Delta t)^2 \cdot \cos(\theta_t) \\ y_{t+1} &= y_t + v_t \cdot \sin(\theta_t) \cdot \Delta t + \frac{1}{2} \cdot a_t \cdot (\Delta t)^2 \cdot \sin(\theta_t) \\ v_{t+1} &= v_t + a_t \cdot \Delta t \\ \theta_{t+1} &= \theta_t + \Delta\theta_t \end{aligned} \tag{5.2}$$

Despite its simplicity, this discrete-time model provides sufficient flexibility for planning and learning in pedestrian-rich environments, where low-speed maneuverability and responsive navigation are prioritized. By omitting mechanical constraints such as steering geometry, it maintains computational efficiency while still capturing the key behavior required for high-level decision-making and safe interaction.

5.3.2 Cognitive Uncertainty Modeling

One of the core computational tasks of the human brain is to infer latent world states and predict future behaviors of others or the environment based on sensory inputs and prior knowledge, especially in uncertain and dynamic conditions [235]. In this process, individuals often experience a lack of full confidence in their predictions. This subjective uncertainty, referred to as cognitive uncertainty [236], reflects how confident an individual is about expected outcomes [237]. When predictions deviate from actual observations, the resulting errors drive belief adjustments and expectation updates, leading to more rational decisions. For example, a pedestrian intending to cross the road may expect an approaching vehicle to slow down and yield. If the pedestrian lacks confidence in this expectation, it reflects a high level of cognitive uncertainty. If the vehicle accelerates instead, the mismatch prompts the pedestrian to re-evaluate and adjust their behavior. Capturing such cognitive processes is vital for modeling human decision-making under uncertainty. Integrating this cognitive mechanism into intelligent agent models allows for more human-like and interpretable behavior, thereby enhancing the realism of human–machine interaction.

The Free Energy Principle (FEP) provides a unifying theoretical framework for describing this cognitive process [238], [239]. It suggests that agents continuously generate predictions and revise internal beliefs based on sensory feedback in order to minimize "surprise" and maintain cognitive stability and behavioral adaptability. In Bayesian Inference, "surprise" is typically formalized as the negative log marginal likelihood, $-\log P(o)$. Since the marginal likelihood $P(o) = \int p(x,o)\,\mathrm{d}x$ is generally intractable, direct minimization is impractical. To overcome this, the FEP introduces variational free energy as an optimizable upper bound [240]:

$$F(Q) = D_{\mathrm{KL}}(Q(x) \| P(x|o)) - \log P(o) \tag{5.3}$$

where $Q(x)$ denotes the agent's approximate posterior belief, and $P(x|o)$ is the true posterior given the observation o. The Kullback-Leibler (KL) divergence quantifies the discrepancy between the two distributions. Since the KL divergence is always non-negative, minimizing free energy pushes the internal belief distribution toward the true posterior. This process provides a mathematical foundation for cognitive modeling.

In practical scenarios, cognitive uncertainty, as a subjective mental state, is inherently difficult to observe or measure directly. Individuals often lack confidence in their judgments when exposed to limited information or complex environments. In order to quantify this uncertainty, we propose the following modeling assumptions: when the discrepancy between predicted and observed outcomes is large, the individual's confidence in

their prediction is low, indicating higher cognitive uncertainty. Accordingly, we adopt the KL divergence between the predicted and the observed distributions as a proxy measure of epistemic uncertainty, capturing the agent's confidence level in forecasting the future behavior of others.

In this work, we use velocity as the target variable and implement a closed-loop cognitive model based on the predict–observe–update process, grounded in the Free Energy Principle and Bayesian Inference. Each agent's prediction of others' velocities is assumed to follow a Gaussian distribution. At each time step t, the ego agent, whether an AV or a pedestrian, generates a velocity prediction $\hat{v}_t$ for surrounding agents, and combines it with the posterior from the previous step $t-1$ to obtain the current prior:

$$p_t(v) = \mathcal{N}(\mu_p, \sigma_p^2) \tag{5.4}$$

Upon observing the actual velocity v_t, an observation distribution is constructed with the observed value as its mean μ_o and a fixed variance σ_o^2:

$$o_t(v) = \mathcal{N}(\mu_o, \sigma_o^2) \tag{5.5}$$

The cognitive uncertainty is approximated by computing the KL divergence between the predicted and observed distributions, expressed as:

$$u_t = D_{KL}(p_t(v)||o_t(v)) \tag{5.6}$$

A larger KL value indicates a greater deviation, implying lower confidence in the prediction and higher epistemic uncertainty. The ego agent then performs Bayesian updating to obtain the posterior [241]:

$$\frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_p^2} + \frac{1}{\sigma_o^2} \tag{5.7}$$

$$\mu_{\text{post}} = \sigma_{\text{post}}^2 \left(\frac{\mu_p}{\sigma_p^2} + \frac{\mu_o}{\sigma_o^2} \right) \tag{5.8}$$

This posterior represents the agent's revised belief about others' behavior at the current time step and serves as the prior for the next prediction, thus forming a recursive cognitive inference process. The complete procedure is presented in Algorithm 4.

Algorithm 4: Human cognitive process modeling with cognitive uncertainty quantification

Input: Surrounding agents set N, time horizon T, internal prediction model
Output: Cognitive uncertainty
for *each time step* $t \in T$ **do**
  for *each surrounding agent* $i \in N$ **do**
    Use internal prediction model to predict velocity $\hat{v}_t^i$;
    Obtain prior distribution $p_t^i(v)$ based on predicted velocity $\hat{v}_t^i$ and posterior $p_{\text{post},t-1}^i(v)$;
    Observe true velocity v_t^i and construct observation distribution $o_t^i(v)$;
    Compute cognitive uncertainty u_t^i;
    Compute posterior distribution $p_{\text{post},t}^i(v)$ using Bayesian update;
    Store the updated posterior $p_{\text{post},t}^i(v)$;

5.3.3 Pedestrian Cognitive-Risk Social Force Model

The social force model [71] describes pedestrian movement as the result of a combination of virtual social forces, including a goal force that drives individuals toward their destinations, and repulsive forces from other agents or obstacles to avoid collisions. These forces are represented as vectors and superimposed to determine the pedestrian's overall motion, enabling modeling and simulation of pedestrian behavior. In the AV–pedestrians interaction considered in this study, each pedestrian is influenced by three primary forces: a goal force guiding movement toward the destination, a vehicle-induced repulsive force from the AV, and repulsive forces from surrounding pedestrians.

To model efficient goal-oriented movement, the goal force for pedestrian i is defined as:

$$\mathbf{F}_i^{\text{goal}} = \frac{v_i^0 \vec{e}_i - \vec{v}_i}{\tau_i} \tag{5.9}$$

where v_i^0 is the desired speed, $\vec{e}_i$ is the unit vector pointing toward the destination, $\vec{v}_i$ is the current velocity of pedestrian i, and τ_i is the relaxation time representing how quickly the pedestrian adapts to the desired velocity. The repulsive interactions are modeled as exponentially decaying functions of distance. The repulsive force from the AV is given by:

$$\mathbf{F}_i^{\text{veh}} = A_{\text{veh}} \cdot \exp\left(\frac{r_{i,\text{veh}} - d_{i,\text{veh}}}{B_{\text{veh}}}\right) \cdot \vec{n}_{i,\text{veh}} \tag{5.10}$$

and the repulsive force from another pedestrian j is computed as:

$$\mathbf{F}_{ij}^{\text{ped}} = A_{ij} \cdot \exp\left(\frac{r_{ij} - d_{ij}}{B_{ij}}\right) \cdot \vec{n}_{ij} \tag{5.11}$$

In both equations, A and B are constants; d is the Euclidean distance between agents; r is the sum of physical radii; and $\vec{n}$ is the unit vector pointing from the interacting agent to pedestrian i. These forces are combined to yield the resultant force acting on each pedestrian. For pedestrian i, the total force can be formulated as [155]:

$$\mathbf{F}_i = \mathbf{F}_i^{\text{goal}} + \mathbf{F}_i^{\text{veh}} + \sum_{\substack{j \in \mathscr{G} \\ j \neq i}} \mathbf{F}_{ij}^{\text{ped}} \tag{5.12}$$

where $\mathscr{G}$ denotes the set of all pedestrians in the current scene.

While the social force model has been widely applied in pedestrian behavior modeling, it still shows certain limitations in complex human-vehicle interaction scenarios. On the one hand, the magnitude of repulsive forces is usually fixed and based solely on spatial distance, without accounting for the velocity and motion direction of interacting agents. For instance, an AV may be physically close to a pedestrian but moving away; in such cases, the risk is negligible. However, the traditional model would still exert a strong repulsive force purely due to proximity, leading to unrealistic or overly cautious avoidance behaviors. On the other hand, the model lacks a representation of human cognitive processes. In real-world interactions, pedestrians continuously perceive, anticipate, and update their beliefs about others' behaviors. This "perception–prediction–observation–update" loop is fundamental to human decision-making in dynamic, uncertain environments. But it is absent in the traditional social force model.

To overcome these limitations, we propose the pedestrian Cognitive-Risk Social Force Model. The core idea is to introduce risk weight coefficients that dynamically modulate the intensity of each force, allowing pedestrian behaviors to adapt responsively to evolving interaction contexts. Fig. 5.2 provides a schematic comparison between the SFM and CR-SFM, showing the difference in resultant force under the same interaction scenario. The revised resultant force acting on pedestrian i is formulated as:

$$\mathbf{F}_i = w_i^{\text{goal}} \cdot \mathbf{F}_i^{\text{goal}} + w_i^{\text{veh}} \cdot \mathbf{F}_i^{\text{veh}} + \sum_{\substack{j \in \mathscr{G} \\ j \neq i}} w_{ij}^{\text{ped}} \cdot \mathbf{F}_{ij}^{\text{ped}} \tag{5.13}$$

where w_i^{goal}, w_i^{veh} and w_{ij}^{ped} represent the risk-based weights for the goal force, vehicle repulsion, and pedestrian repulsion, respectively. These weights are computed by combining physical risk and cognitive uncertainty.

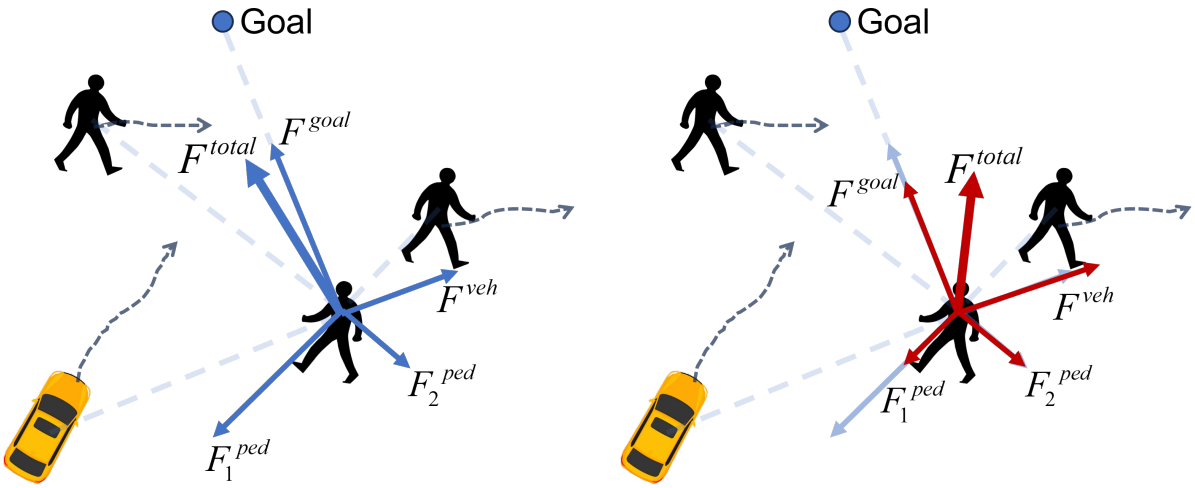


Figure 5.2: Comparison of the SFM and CR-SFM in pedestrian-vehicle interactions. These figures show the force composition acting on the ego pedestrian under the SFM (a) and CR-SFM (b). In the SFM, all repulsive forces are distance-based with fixed weights. Pedestrian 1 (the rightmost pedestrian in the figure), being closest, produces the strongest repulsion, shifting the total force toward the AV and increasing collision risk. In CR-SFM, force weights are adjusted based on physical features (e.g., distance, velocity, direction) and cognitive uncertainty. Although pedestrian 1 is nearest, his motion away leads to low perceived risk and a weaker force. The approaching AV contributes more strongly, shifting the total force toward a safer direction.

In dynamic interaction scenarios, risk assessment methods based solely on geometric distance often fail to accurately capture the actual level of threat, as they overlook the critical influence of relative motion on risk perception. For example, at the same distance, an agent approaching at high speed poses a greater danger than one moving slowly or moving away. To better reflect this perceptual difference, we introduce a risk modeling approach based on virtual distance. By incorporating dynamic motion features such as velocity, acceleration, and direction [222] to modulate the actual distance, the resulting virtual distance is then used to replace the original geometric distance for physical risk computation. The physical risk ψ is defined as:

$$d_{\text{virtual}} = d_{\text{actual}} \cdot (1 + \tanh(\gamma_1 \cdot k \cdot (v \cdot |\cos\varphi| + |a|))) \tag{5.14}$$

$$k = \begin{cases} 1, & \text{if the agent is moving away} \\ -1, & \text{if the agent is approaching} \end{cases} \tag{5.15}$$

$$\psi = \frac{1}{1 + \gamma_2 \cdot d_{\text{virtual}}} \tag{5.16}$$

where d_{actual} is the actual physical distance between the pedestrian and the interacting agent, v and a are the velocity and acceleration of the interacting agent, φ is the angle between the agent's velocity vector and the vector pointing from the agent to pedestrian i, and γ_1 and γ_2 are hyperparameters. These formulations imply that risk increases when an approaching agent moves towards the pedestrian with higher speed or greater acceleration.

Cognitive uncertainty captures the pedestrian's subjective confidence in predicting others' future behaviors. Rather than being modeled as an independent risk component, it is treated as a modulatory factor that amplifies physical risk under uncertain situations. For example, when physical risk is low, but the pedestrian is unsure of the AV's future behavior, a higher cognitive uncertainty increases perceived risk, enhancing repulsive forces and encouraging earlier avoidance. The risk weight coefficients are calculated as follows:

$$w_i^{\text{veh}} = \psi_i^{\text{veh}} \cdot \left(1 + \lambda_1 \cdot u_i^{\text{veh}}\right) \tag{5.17}$$

$$w_{ij}^{\text{ped}} = \psi_{ij}^{\text{ped}} \cdot \left(1 + \lambda_2 \cdot u_{ij}^{\text{ped}}\right) \tag{5.18}$$

$$w_i^{\text{goal}} = \exp\left(-\lambda_3 \cdot \max\left(w_i^{\text{veh}}, \max_{j \in \mathscr{G}, j \neq i} w_{ij}^{\text{ped}}\right)\right) \tag{5.19}$$

where u_i^{veh} and u_{ij}^{ped} denote cognitive uncertainties by the AV and by surrounding pedestrians, respectively. λ_1, λ_2, λ_3 are scaling hyperparameters. Notably, the goal force weight is modeled using an exponential decay function to reduce goal-oriented behavior when perceived interaction risk is high, allowing the pedestrian to prioritize avoidance over forward motion.

In summary, by introducing a dynamic weighting mechanism that integrates both physical and cognitive factors, the CR-SFM can make the pedestrian to modify their approach and avoidance strategies based on perceived risk and uncertainty, more closely aligning with real human decision-making processes.

5.3.4 Graph-Enhanced DRL for AV Decision-Making

In this study, we adopt the SAC as a basic reinforcement learning framework to guide the AV decision-making strategies in complex, pedestrian-rich environments. SAC is employed due to its suitability for uncertain and interactive AV-pedestrian environments. Its entropy-regularized objective enables effective exploration, and its clipped double-Q structure reduces Q-value overestimation, improving training stability. As an off-policy algorithm, SAC is also more sample-efficient than on-policy alternatives. In contrast, DDPG's deterministic policy limits exploration, and PPO suffers from lower sample ef-

ficiency, making them less suitable for this task. The SAC is grounded in the Maximum Entropy Reinforcement Learning theory. While optimizing the expected cumulative return, it explicitly introduces a policy entropy term to encourage greater stochasticity and exploration under uncertainty. The objective is to maximize the following expression [242]:

$$J(\pi) = \sum_t \mathbb{E}_{(s_t,a_t)\sim\pi}\left[r(s_t,a_t) + \eta \mathscr{H}(\pi(\cdot|s_t))\right] \tag{5.20}$$

where $\mathscr{H}(\pi(\cdot|s_t))$ is the entropy of the policy at state s_t, and η is a temperature coefficient balancing reward and entropy.

The SAC follows an actor-critic architecture consisting of a policy network (actor), two Q-value networks (critics), and their corresponding target networks. The actor outputs a Gaussian distribution over actions given a state s_t, from which an action a_t is sampled. The Q-networks estimate the value of the state-action pair (s_t, a_t). Using samples from a replay buffer, the networks are updated iteratively to approximate optimal behavior. To mitigate Q-value overestimation, SAC employs a double Q-network structure, where the minimum of the two Q-values is used to compute the target value. The critic is trained to minimize the following loss:

$$\mathscr{L}_{\text{critic}} = \mathbb{E}_{(s_t,a_t,r_t,s_{t+1})\sim\mathscr{D}}\left[\frac{1}{2}\left(Q(s_t,a_t) - y_t\right)^2\right] \tag{5.21}$$

with the target value y_t defined as:

$$y_t = r_t + \gamma \min_{i=1,2} Q_i'(s_{t+1}, a_{t+1}) - \eta \log \pi(a_{t+1} \mid s_{t+1}) \tag{5.22}$$

The target Q-networks are updated softly via:

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta' \tag{5.23}$$

where $\tau \in (0,1)$ is a small soft update coefficient, θ denotes the parameters of the current critic network, and θ' denotes the parameters of the target critic network. The actor is trained to maximize the expected Q-value while maintaining high entropy:

$$\mathscr{L}_{\text{actor}} = \mathbb{E}_{s_t\sim\mathscr{D}}\left[\eta \log \pi(a_t|s_t) - Q(s_t,a_t)\right] \tag{5.24}$$

Through this maximum entropy optimization framework, the SAC enables stable and continuous policy learning in complex environments, allowing the AV to develop strategies that are both efficient and exploratory in dynamic multi-agent interactions.

However, the standard SAC architecture does not explicitly model dynamic interactions among multiple agents, which is critical in autonomous driving scenarios involving dense pedestrian. To address this, we incorporate a GCN into the SAC framework to extract interaction-aware state representations from the multi-agent graph.

At each time step, the AV and surrounding pedestrians are modeled as graph nodes, with the adjacency matrix encoding their interactions. Given a graph $G_t = (V_t, A_t)$, where V_t is a set of all agents at time t, A_t is the adjacency matrix, $H^{(l)} \in \mathbb{R}^{N \times d}$ represents the node feature at layer l, and the GCN propagation rule is expressed as [243]:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H^{(l)}W^{(l)}\right) \tag{5.25}$$

where $\tilde{A} = A + I$ is the adjacency matrix adding self-connection, $\tilde{D}$ is the degree matrix, $W^{(l)}$ is the learnable weight matrix, and σ is the nonlinear activation function. This operation aggregates each agent's state with information from its neighbors, capturing local structural relationships and interaction patterns. The resulting high-order representations serve as enhanced state features for downstream policy learning and value estimation.

However, the adjacency matrix in conventional GCNs is often constructed using static graph topology, making it insufficient for modeling dynamic and asymmetric interaction patterns over time. In real-world multi-agent interactions, humans do not pay equal attention to all agents, but instead focus on potentially high-risk individuals that are closer, faster, or have unpredictable behaviors.

Inspired by this, we propose a risk encoded interaction graph, where physical risk and cognitive uncertainty are combined into a unified risk score. This value directly determines the attention weights in the adjacency matrix, allowing the model to focus more effectively on high-risk interactions during information aggregation. Specifically, we construct a dynamic adjacency matrix A_t at each time step, where each element $A_t(i, j)$ represents the influence of agent j on agent i:

$$A_t(i, j) = \psi_t(i, j) \cdot (1 + \lambda \cdot u_t(i, j)) \tag{5.26}$$

where $\psi_t(i, j)$ denotes the physical risk based on agent j's relative position, speed, and direction with respect to agent i, while $u_t(i, j)$ quantifies agent i's uncertainty about agent j's behavior, computed using the free-energy-based method described in Algorithm 4. λ is a hyperparameter controlling the influence of uncertainty.

Algorithm 5: SAC for AV Decision-Making with Cognitive Process.

Initialize critic networks Q_1, Q_2, actor network π, target networks Q_1', Q_2', and replay buffer $\mathscr{D}$;

for *episode* $i \in M$ **do**

  Observe initial state s_0;

  for *each time step* $t \in T$ **do**

    Compute cognitive uncertainty u_t using Algorithm 4;

    Execute an action a_t, receive a reward r_t, and observe the next state s_{t+1};

    Store the transition $(s_t, a_t, r_t, s_{t+1}, u_t)$ in replay buffer $\mathscr{D}$;

    Update $s_t \leftarrow s_{t+1}$;

    Sample a minibatch of transitions from $\mathscr{D}$;

    Compute target Q-value y_t;

    Update critics Q_1 and Q_2 by minimizing loss $\mathscr{L}_{\text{critic}}$;

    Update actor network π;

    Soft update target critic networks Q_1' and Q_2';

This dynamic adjacency matrix guides GCN propagation to focus on high-risk interactions, enabling the model to emphasize critical agents and filter out irrelevant ones. Compared to traditional graph modeling, our approach jointly encodes physical structure and cognitive cues, providing the policy network with more decision-relevant state representations. This enhances the policy's ability to identify key interactions and make robust decisions in complex, multi-agent environments. The overall training process is summarized in Algorithm 5.

5.3.5 Reward Function Formulation

The design of the reward function is a critical component of RL algorithms, as it directly guides the agent's behavior during training. To encourage the AV to exhibit safe, efficient, and socially compliant behaviors in pedestrian-rich environments, we construct a composite reward function that integrates multiple objectives. At each timestep, the total reward r_{total} is computed as a weighted sum of several components, each capturing a distinct aspect of the AV's driving performance:

$$r_{\text{total}} = r_{\text{success}} + r_{\text{collision}} + r_{\text{efficiency}} + r_{\text{direction}} + r_{\text{speed}} + r_{\text{comfort}} + r_{\text{step}} \tag{5.27}$$

A goal achievement reward $r_{success}$ is assigned to encourage task completion. A positive reward is given when the AV reaches its predefined destination:

$$r_{\text{success}} = \begin{cases} 30, & \text{if goal reached} \\ 0, & \text{otherwise} \end{cases} \tag{5.28}$$

To penalize unsafe behavior, a collision penalty $r_{collision}$ is imposed if the AV collides with any pedestrian:

$$r_{\text{collision}} = \begin{cases} -30, & \text{if collision with pedestrians} \\ 0, & \text{otherwise} \end{cases} \tag{5.29}$$

To guide the AV toward the goal efficiently and along a correct heading, we design a goal-directed efficiency reward. This component rewards reductions in distance to the destination, modulated by a direction alignment factor:

$$r_{\text{efficiency}} = 2 \cdot (d_{\text{goal}}^{t-1} - d_{\text{goal}}^{t}) \cdot \max\left(0, 1 - \frac{\Delta\theta_{\text{dev}}}{0.1}\right) \tag{5.30}$$

where d_{goal}^{t-1} and d_{goal}^{t} represent the distance from the agent's previous and current positions to the goal position, respectively. $\Delta\theta_{\text{dev}}$ denotes the directional deviation between the AV's current velocity vector and the vector pointing towards the goal. In addition, a direction penalty is applied to discourage large angular deviations from the goal trajectory:

$$r_{\text{direction}} = -2 \cdot \Delta\theta_{\text{dev}} \tag{5.31}$$

During training, the agent often encounters the freezing problem, a common form of local optimum in which the agent chooses to remain stationary in response to nearby risks. While this behavior may reduce immediate danger, it can lead to long-term inefficiency and goal failure, as the agent continues to stay immobile even after the risk has subsided. To mitigate such overly conservative behavior, a speed penalty is introduced when the AV's velocity drops below a predefined threshold of 0.6 m/s. This penalty encourages the agent to maintain a minimal level of motion, thereby avoiding stagnation and fostering more proactive and goal-directed behavior. The corresponding reward r_{speed} is defined as:

$$r_{\text{speed}} = -3 \cdot \max\left(0, 1 - \frac{v_t}{0.6}\right) \tag{5.32}$$

Passenger comfort is considered by penalizing abrupt acceleration changes, encouraging smooth driving:

$$r_{\text{comfort}} = -0.5 \cdot |a_t - a_{t-1}| \tag{5.33}$$

A small constant step penalty r_{step} (-0.5) is added at every timestep to motivate the AV to complete the task in fewer steps. Together, these reward terms provide dense feedback across safety, efficiency, and comfort dimensions, supporting the AV in learning robust driving strategies.

5.4 Experiment

5.4.1 Experimental Setup

To evaluate the effectiveness of the proposed pedestrian CR-SFM and the AV decision-making model, we conducted a series of simulation experiments in a shared-space interaction scenario.

We adopted the real-world Hamburg Bergedorf Station (HBS) dataset [10], which captures a typical shared-space street environment in Hamburg, Germany. This dataset contains rich interactions between low-speed vehicles and multiple crossing pedestrians in unconstrained urban settings. From the HBS dataset, we extracted 78 representative interaction scenarios, each involving a single AV and three pedestrians. These scenarios were divided into a training set (77%) and test set (23%) to support model learning and performance evaluation, respectively. Each simulation was initialized using the recorded initial positions and velocities of all agents. During the interaction process, the AV generated its actions based on the learned policy, while pedestrian behaviors were governed by our proposed CR-SFM. A simulated episode terminated when the AV reached its goal, collided with a pedestrian, or the maximum time horizon was exceeded.

All experiments were conducted on a high-performance laptop equipped with an Intel i9 processor and an NVIDIA RTX 4090 GPU. Model training and simulation were implemented using PyTorch in a custom-built environment. Due to computational limitations on local hardware, we restricted our simulation setup to include three pedestrians interacting with an AV. This configuration balances training feasibility and interaction complexity, while still capturing key social dynamics observed in shared-space environments. In the simulation, the AV's maximum speed was set to 6 m/s, with acceleration constrained to the range of [-2, 2] m/s^2. Pedestrians were limited to a maximum walking speed of 2 m/s. The simulation operated with a discrete time step of 0.5 seconds. The detailed hyperparameter configuration of the SAC framework is provided in Table 5.1. Particularly, the actor and critic networks share the same architecture, comprising one GCN layer (with hidden dimension 16) and three MLP layers (each with 256 hidden units). The temperature coefficient for entropy regularization is automatically tuned, initialized at 0.2.

Table 5.1: Hyperparameter settings of the SAC framework.

Parameter	**Description**	**Value**
Buffer capacity	Maximum number of transitions stored	100000
Batch size	Number of samples per update	256
Discount factor γ	Future reward discount	0.99
Soft update rate τ	Target network update rate	0.005
Learning rate (actor)	Learning rate for actor network	5×10^{-5}
Learning rate (critic)	Learning rate for critic network	5×10^{-5}
Learning rate (entropy temperature)	Learning rate for entropy temperature coefficient	1×10^{-5}
Initial entropy coefficient η	Initial value for entropy temperature parameter	0.2
Target entropy	Desired target entropy value	-2
Max steps per episode	Maximum time steps per episode	50
Update interval	Policy update interval in steps	32

5.4.2 Comparative Models and Evaluation Metrics

This section describes the comparative models as well as evaluation indicators used for both the pedestrian model and the AV decision-making model.

For the pedestrian model, in order to assess the performance of our proposed pedestrian CR-SFM, we conducted comparative experiments with two baseline models and one ablation variant. The compared models are detailed as follows:

Constant Velocity Model (CV) [12]: In this model, the pedestrian maintains a constant initial velocity toward the predefined goal throughout the entire episode. The motion is purely goal-driven and remains unaffected by the presence or behavior of surrounding agents.

Social Force Model (SFM) [155]: This classical model serves as the primary baseline. It describes the pedestrian's motion as the result of three interacting forces, as defined in Equation 5.12. In the standard formulation, the magnitudes of these forces are uniformly weighted using fixed coefficients (set to 1), without any dynamic modulation in response to changes in interaction context.

Risk-aware Social Force Model (RA-SFM): This variant serves as an ablation model to evaluate the specific contribution of cognitive uncertainty. It extends the standard SFM by introducing physical risk as a modulation factor for social force magnitudes. Specifically, the strengths of the goal force, vehicle repulsion force, and pedestrian repulsion force are dynamically adjusted based on the estimated physical risk from surrounding agents.

Cognitive-Risk Social Force Model (CR-SFM): This is our proposed model. Building upon the RA-SFM, CR-SFM incorporates both physical risk and cognitive uncertainty into the modulation of force magnitudes.

To ensure a fair comparison, all models are calibrated using the same set of real-world pedestrian trajectories extracted from the training dataset. Bayesian Optimization is employed to fit the model parameters, aiming to best replicate the real pedestrian behaviors. All models are operated under the same experimental conditions, using the following three metrics for trajectory evaluation:

Average Displacement Error (ADE): ADE measures the average deviation between the simulated trajectory and the ground truth trajectory across all time steps. For each episode, we compute the mean Euclidean distance between simulated and actual positions across every time step. The final ADE is obtained by averaging these values over all episodes [244]:

$$\text{ADE} = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{x}_t - x_t\|\right) \tag{5.34}$$

where M is the total number of test episodes, T is the number of time steps in each episode, $\hat{x}_t$ denotes the simulated position of at time step t, and x_t is the corresponding ground-truth position.

Final Displacement Error (FDE): FDE evaluates the final-point accuracy by measuring the Euclidean distance between the simulated final position and the true final position [244]:

$$\text{FDE} = \frac{1}{M}\sum_{m=1}^{M}\|\hat{x}_T - x_T\| \tag{5.35}$$

Collision Rate (CR): CR is defined as the proportion of simulation episodes in which the simulated pedestrian model collides with other agents. It is calculated as the ratio of collision cases to the total number of episodes. A lower CR value indicates better collision avoidance performance.

To evaluated our AV decision-making model, we compare it with one strong baseline and two ablation variants:

Uncertainty-Aware Polar Collision Grid (UAW-PCG) [13]: This state-of-the-art DRL framework integrates pedestrian trajectory prediction uncertainty into AV navigation in crowded environments. It employs a Polar Collision Grid predictor to estimate future pedestrian positions along with their associated uncertainty. These predictions are then incorporated into both the input state and the design of the reward function. The AV decision-making policy is optimized using Proximal Policy Optimization (PPO) algorithm.

Standard SAC (S-SAC): This baseline uses a standard SAC framework, where both the actor and critic networks consist only of multilayer perceptron without GCN modules. It does not explicitly capture interactions between the AV and surrounding agents, and serves to evaluate the benefit of introducing graph-based structures.

Graph-enhanced SAC without cognitive modeling (G-SAC-NoCog): This ablation model adopts the GCN module into the SAC framework to enhance the perception of multi-agent interactions. However, it uses a static adjacency matrix with uniform weights between all agent pairs, ignoring both physical risk and cognitive uncertainty. This model is designed to isolate the contribution of our proposed risk-encoded interaction representation.

Graph-enhanced SAC with cognitive modeling (G-SAC-Cog): This is our proposed model. It integrates a GCN module to capture structured multi-agent interactions, with the adjacency matrix dynamically constructed based on a fused risk signal that combines physical risk and cognitive uncertainty.

Following the evaluation protocol used in [13], [245], we assess each model from the perspectives of safety, efficiency and smoothness. Specifically, safety is quantified using three indicators: the success rate (the proportion of episodes in which the AV reaches its destination without any collisions), the collision rate (the frequency of collisions with surrounding agents), and the timeout rate, which accounts for episodes where the AV fails to reach the goal within the maximum allowed time despite avoiding collisions. Efficiency is evaluated based on the average vehicle speed, which reflects the efficiency of the AV's navigation performance. Lastly, smoothness is measured by the average jerk and average maximum absolute acceleration/deceleration, capturing the stability and comfort of the AV's motion during navigation.

5.5 Results and Discussion

This section presents the qualitative and quantitative validation results of both the proposed pedestrian CR-SFM and AV decision-making model.

5.5.1 Pedestrian Cognitive-Risk Social Force Model

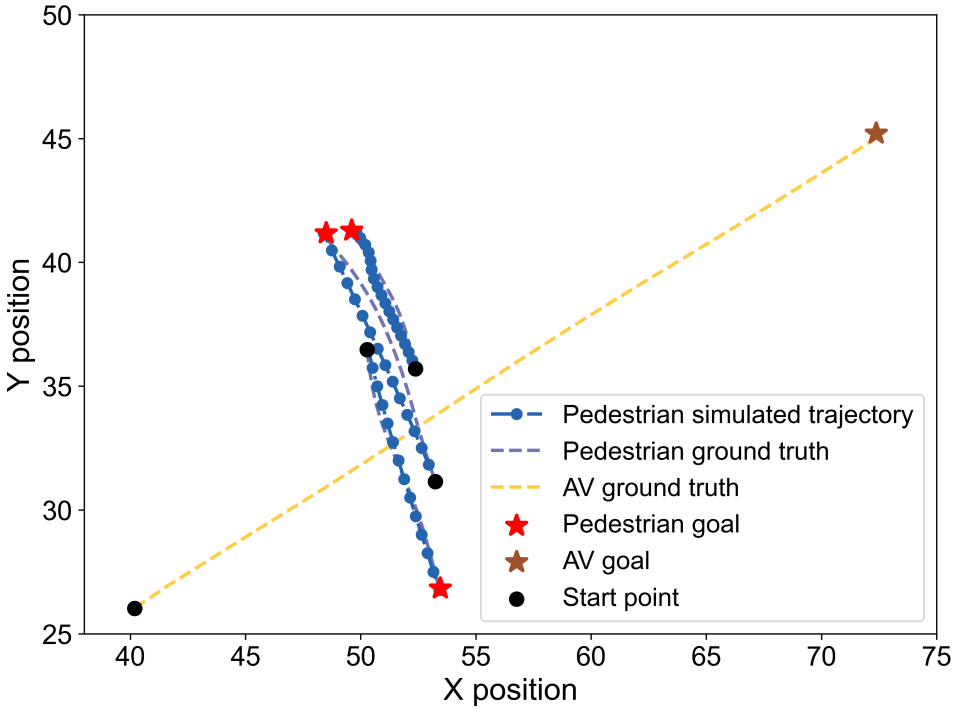
5.5.1.1 Qualitative Analysis

To demonstrate how the proposed pedestrian model performs under different interaction conditions, three representative examples are presented in Fig. 5.3.

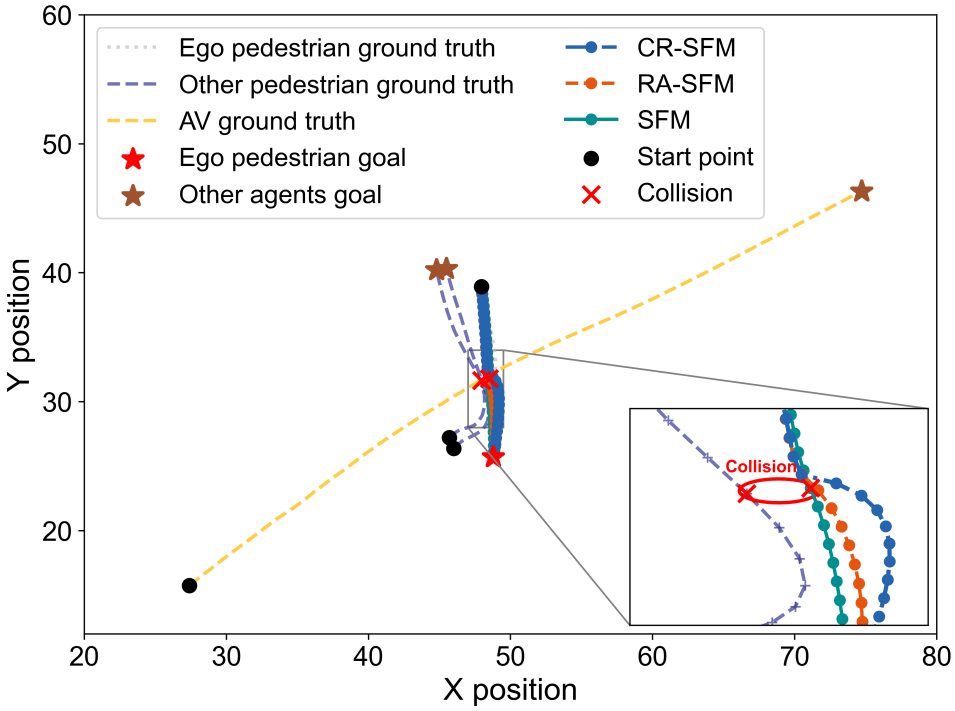
Fig. 5.3a illustrates a normal case in which the AV follows its ground-truth trajectory, while pedestrians are simulated using the proposed CR-SFM. As shown in the figure, all pedestrians safely reach their respective destinations without any collisions. The simulated trajectories (blue triangles) show high similarity to the ground-truth pedestrian trajectories (purple crosses), with only minor deviations. This confirms that the model's ability to generate safe and goal-directed motion in interactive scenarios.

Fig. 5.3b provides a scenario involving an explicit collision risk between the simulated pedestrian and a nearby agent. In this case, the AV and two pedestrians follow their ground-truth trajectories, while the third pedestrian's trajectories are simulated by the SFM, RA-SFM, and CR-SFM. The zoomed-in subfigure reveals that the SFM-generated trajectory leads to a collision with another pedestrian, with collision point marked by a red cross. In contrast, both the RA-SFM and CR-SFM successfully avoid the hazard by slightly deviating the pedestrian's path away from the surrounding agent, toward the right-hand side of the scene. This behavioral divergence highlights the benefit of adaptive social force modulation in RA-SFM and CR-SFM, enabling the agent to respond effectively to immediate physical threats. Notably, the trajectory generated by the CR-SFM displays a more pronounced shift toward the safer region, indicating a higher sensitivity to dynamic interaction contexts.

Fig. 5.3c further investigates the model's behavior under ambiguous or latent risk conditions. Although no immediate physical danger is apparent, both the SFM and RA-SFM trajectories continue directly toward the goal and result in collisions, as shown in the zoomed-in area. In contrast, the CR-SFM trajectory shows a markedly different pattern: the pedestrian perceives elevated risk in the goal direction, detours toward a low-risk zone, and resumes goal-directed motion only after the uncertain agent has passed.

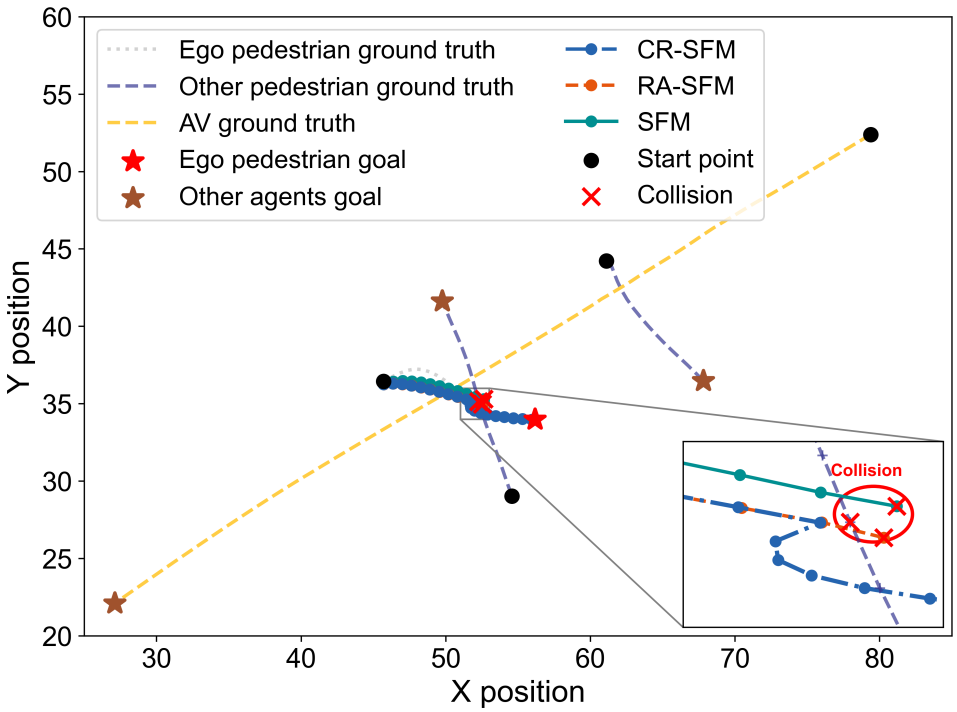


(a)



(b)

Figure 5.3: Qualitative examples of pedestrian trajectory simulation under different interaction scenarios. (a) A normal case in which all pedestrian trajectories are simulated by the proposed CR-SFM. (b) A comparison of three trajectories generated by the SFM, RA-SFM, and CR-SFM. The trajectory generated by the SFM results in a collision, while the RA-SFM and CR-SFM avoid the nearby agent.



(c)

Figure 5.3: (Continued) Qualitative examples of pedestrian trajectory simulation under different interaction scenarios. (c) The SFM and RA-SFM fail to trigger avoidance behavior, whereas the CR-SFM proactively detours before reaching the goal.

These behavioral differences stem from the models' risk perception capabilities. The SFM relies on fixed interaction strengths based purely on distance, and thus cannot adapt to changing risk profiles. RA-SFM improves on this by adjusting repulsive forces according to observable physical cues such as proximity and velocity. However, when these cues remain below a critical threshold, the model fails to trigger sufficient repulsion, and the pedestrian proceeds along a collision-prone path without taking an evasive action.

In contrast, the CR-SFM introduces cognitive uncertainty modeling, allowing the pedestrian to account not only for physical risk but also for uncertainty in the predicted behaviors of nearby agents. By quantifying the discrepancy between the expected and observed motion, CR-SFM captures the subjective uncertainty and then integrates it into the overall risk assessment. This allows the model to detect latent threats that are not evident through physical cues alone. Consequently, even when the physical risk is low, the elevated uncertainty can amplify the perceived danger and increase the total risk. When this combined risk becomes significant, the pedestrian proactively adjusts its trajectory, even temporarily moving away from the destination to avoid potential collisions. Once the perceived risk diminishes, the pedestrian returns to the original path and safely reaches the goal. This behavior illustrates the unique advantage of cognitive uncertainty modeling in CR-SFM. By integrating internal uncertainty estimation with external physical

observations, the model is able to anticipate possible conflicts and make more informed decisions. As a result, CR-SFM generates more flexible, cautious, and human-like behavior in complex multi-agent environments, significantly enhancing both trajectory safety and realism.

In conclusion, the CR-SFM demonstrates strong performance for realistic pedestrian simulation, enabling reliable behavior modeling under diverse and challenging conditions.

5.5.1.2 Quantitative Analysis

Table 5.2 summarizes the quantitative performance of various pedestrian models in terms of ADE, FDE, and CR. The CV model yields the poorest performance, with an ADE of 1.5262 and an FDE of 2.8738, indicating large deviations from the ground-truth trajectories. It also records the highest collision rate, which is expected given that the CV model neglects goals, interactions, or environmental dynamics. In contrast, the SFM introduces interactive forces and goal attraction, significantly lowering the ADE and FED to 0.8310 and 0.6151, respectively. However, it still suffers from a non-negligible collision rate (3.7%), highlighting its limitation in complex interactions due to its reliance on fixed interaction parameters.

Table 5.2: Statistic results of our pedestrian model compared with other approaches.

Pedestrian Model	ADE	FDE	CR
CV [12]	1.5262	2.8738	0.0926
SFM [155]	0.8310	0.6151	0.0370
RA-SFM	0.8139	0.5247	0.0370
CR-SFM (Ours)	**0.7842**	**0.4416**	**0.0**

The RA-SFM enhances the traditional SFM by making the repulsive force adaptive to kinematic features such as distance, velocity and direction. This results in a slight improvement in the simulation accuracy, but the collision rate remains unchanged. Although RA-SFM allows for more adaptive motion generation, it still lacks the capacity to account for uncertainty in agent behaviors.By comparison, the proposed CR-SFM outperforms all baselines across all metrics. It obtains the lowest ADE (0.7842) and FDE (0.4416), reflecting its superior trajectory simulation accuracy. Notably, it achieves a zero collision rate, successfully avoiding all collisions in the evaluation scenarios. This performance is primarily attributed to the integration of cognitive risk modeling, which enables the pedestrian to not only react to visible physical cues but also anticipate potential conflicts arising from uncertainty in others' behaviors.

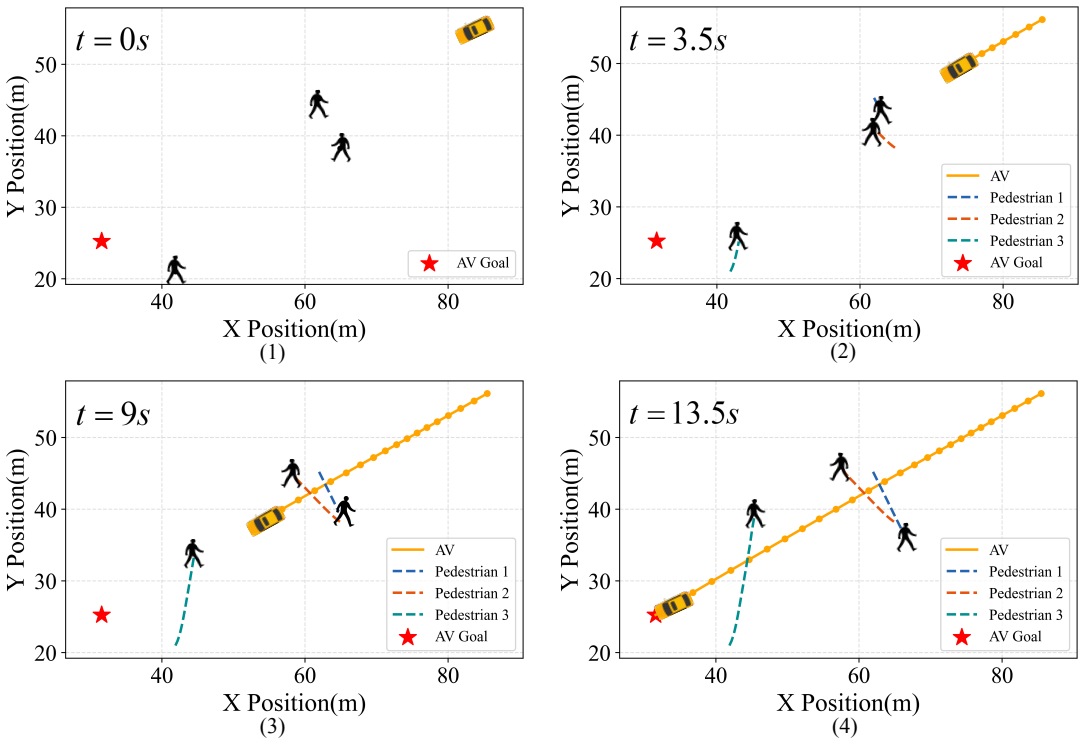
In summary, the quantitative results demonstrate that the CR-SFM achieves reliable performance in both accuracy and safety. Compared to other methods, it reduces displacement errors while enhancing collision avoidance. These findings indicate that combining cognitive risk reasoning with adaptive social forces enables more flexible and robust generation of pedestrian behavior in dynamic and interactive environments.

5.5.2 AV Decision-Making Model

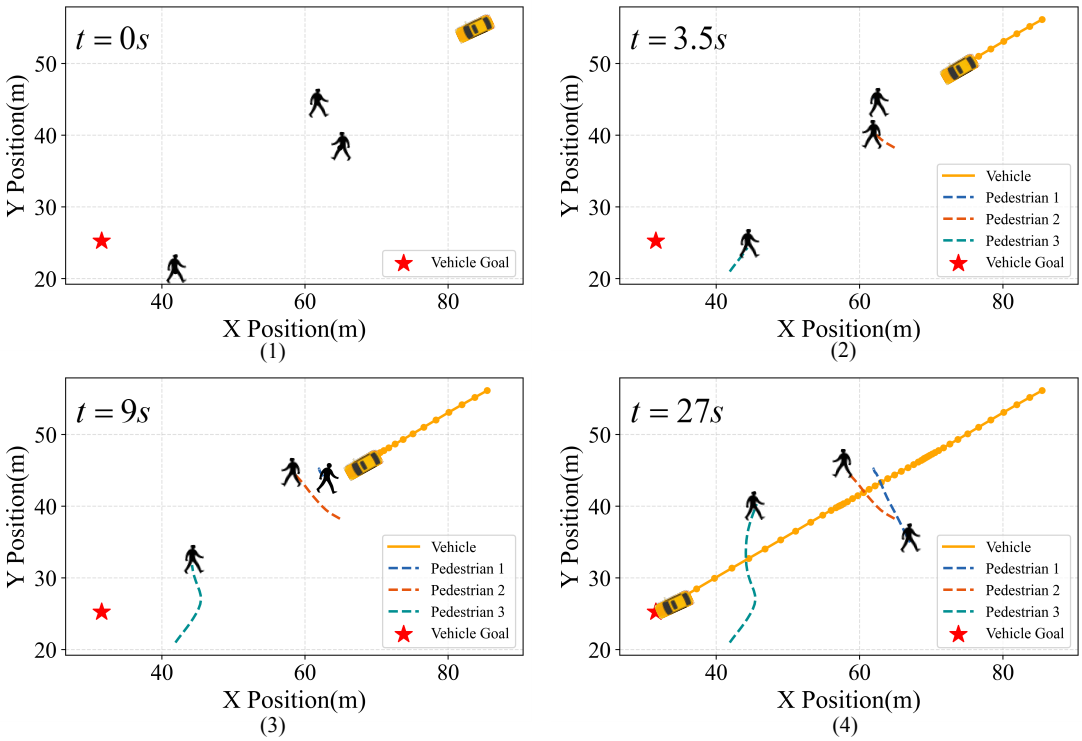
5.5.2.1 Qualitative Analysis

To clearly demonstrate how our AV decision-making model operates across different scenarios, three representative cases are presented below.

Case 1: provides an example where both the AV and the human driver decelerate early to avoid collisions with nearby pedestrians, as shown in Fig. 5.4. At the beginning of the episode (Fig. 5.4a and Fig. 5.4b, top-left), the AV is at its initial position with three pedestrians distributed across the scene. Pedestrian 1 stands particularly close to the AV's planned path, indicating a higher potential collision risk. In the simulation (Fig. 5.4a), the AV gradually decelerates in response to the pedestrians' anticipated behavior, slowing sufficiently to safely pass behind Pedestrians 1 and 2 before smoothly accelerating again once the path is clear. During the interaction with Pedestrian 3 (Fig. 5.4a, bottom-left), the AV maintains a low, steady acceleration without significant speed reduction, reflecting its assessment of minimal risk and resulting in a smooth and continuous motion strategy. Finally, the AV successfully navigates through all pedestrians and reaches its goal without abrupt stops or collisions.

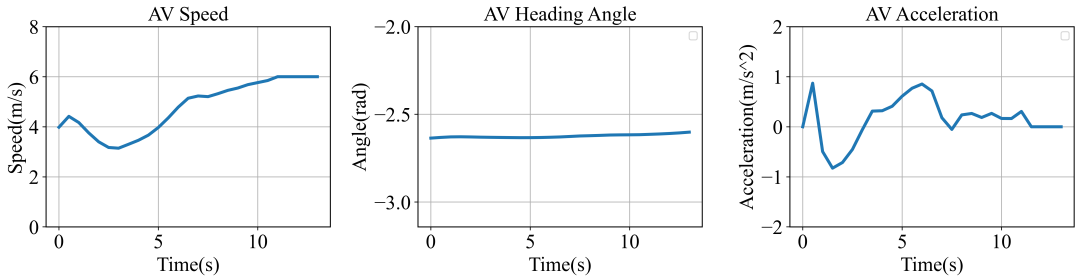


(a)

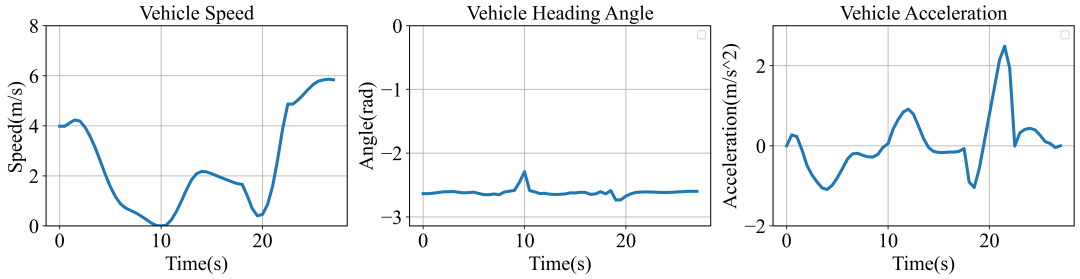


(b)

Figure 5.4: Comparison of the interaction process and dynamic state evolution between the proposed model and real driving data for Case 1. (a) Interaction simulation using the proposed model. (1) State at t = 0 s: v_{AV} = 3.981 m/s, θ = -2.635 rad, a_{AV} = 0 m/s^2 (2) State at t = 3.5 s: v_{AV} = 3.144 m/s, θ = -2.630 rad, a_{AV} = -0.057 m/s^2 (3) State at t = 9 s: v_{AV} = 5.455 m/s, θ = -2.619 rad, a_{AV} = 0.265 m/s^2 (4) State at t = 13.5 s: v_{AV} = 6.0 m/s, θ = -2.601 rad, a_{AV} = 0 m/s^2. (b) Interaction process captured from real driving data. (1) State at t = 0 s: v_{veh} = 3.981 m/s, θ = -2.635 rad, a_{veh} = 0 m/s^2. (2) State at t = 3.5 s: v_{veh} = 3.578 m/s, θ = -2.601 rad, a_{veh} = -0.903 m/s^2. (3) State at t = 9 s: v_{veh} = 0.266 m/s, θ = -2.597 rad, a_{veh} = -0.281 m/s^2. (4) State at t = 27 s: v_{veh} = 5.838 m/s, θ = -2.599 rad, a_{veh} = 0 m/s^2.



(c)



(d)

Figure 5.4: (Continued) Comparison of the interaction process and dynamic state evolution between the proposed model and real driving data for Case 1. (c) Evolution of AV states predicted by the proposed model. (d) Evolution of vehicle states recorded from real driving data.

The pedestrian trajectories in the simulation (Fig. 5.4a) are generated by our previously proposed pedestrian CR-SFM model, whereas those in Fig. 5.4b are recorded from actual pedestrian movements. In the real-world scenario, although Pedestrian 1 chooses to stop and yield, the human driver also decelerates to a full stop, waiting until the pedestrian has completely crossed before resuming movement. After passing Pedestrians 1 and 2, the driver again reduces speed significantly to yield when approaching Pedestrian 3. These conservative actions lead to marked fluctuations in the vehicle's speed and acceleration profiles, as illustrated in Fig. 5.4d. In contrast, our proposed model achieves smoother and more stable driving behavior by applying sufficient deceleration without unnecessary stops.

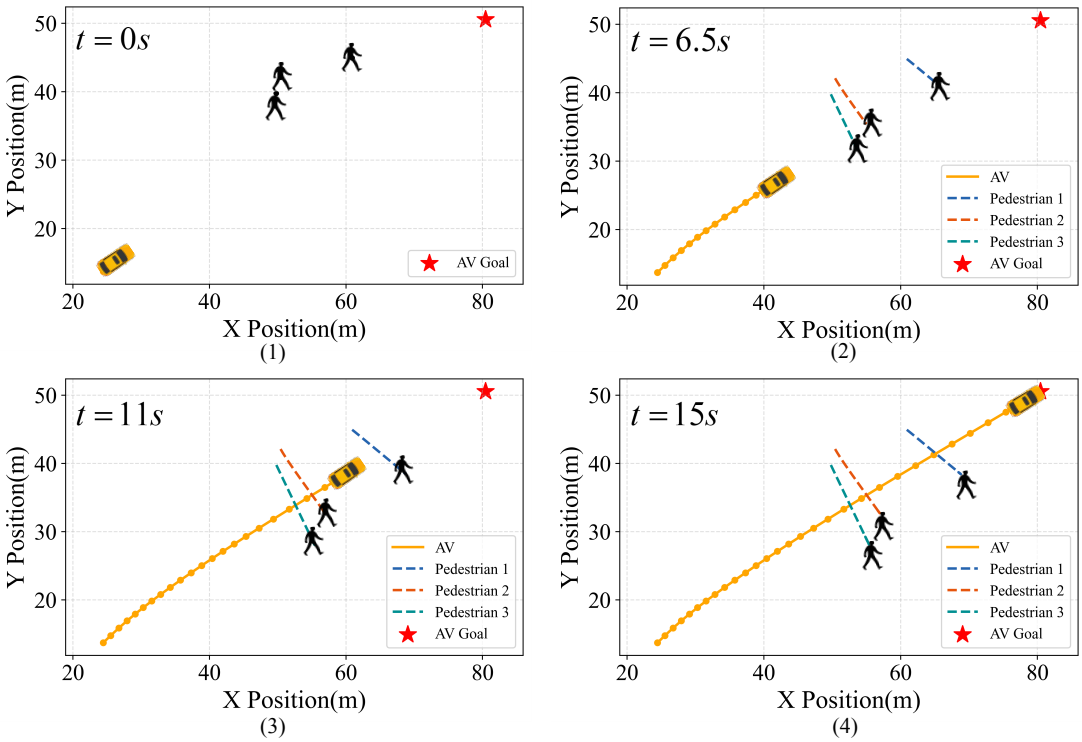
Case 2: illustrates a situation in which both the AV and the driver maintain their speed to yield to pedestrians initially positioned at a greater distance, as shown in Fig. 5.5. Early in this episode, both the AV and the human driver perceive that the pedestrians will enter their intended path. However, given the sufficient distance, they evaluate the situation as low-risk and opt to maintain a speed of approximately 3.5 m/s without significant braking. This behavior contrasts with Case 1, where the closer proximity of pedestrians required early deceleration to ensure safety.

Once the pedestrians have safely crossed and the potential risk is eliminated, the human driver continues at a steady speed of around 4 m/s, whereas the AV accelerates smoothly to its maximum velocity to efficiently reach its goal. Throughout the interaction, both the AV and the driver exhibit similar decision-making patterns, with consistent trends observed in their acceleration profiles and heading angle adjustments (Fig. 5.5c and Fig. 5.5d). These results suggest that the overall behavior of the proposed model closely mirrors that of the human driver, while showing smoother and more stable heading angle transitions.

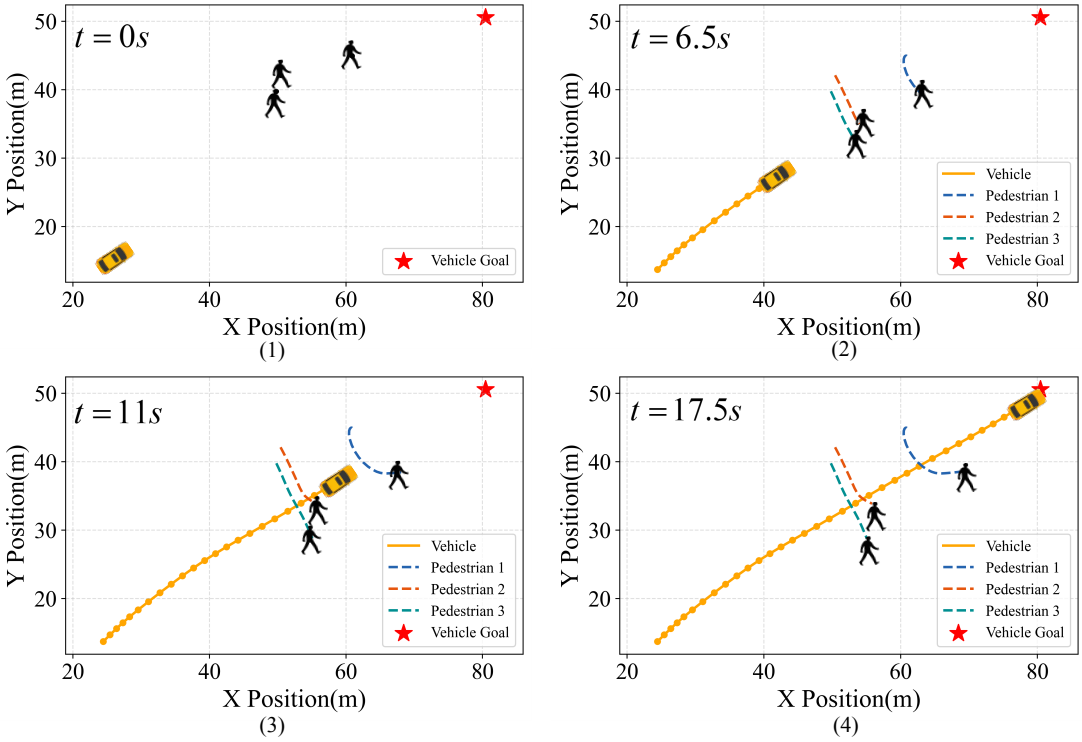
Case 3: illustrates a scenario in which both the AV and a human driver exhibit significant deceleration to yield to pedestrians, as illustrated in Fig. 5.6. At the beginning of the episode, the nearest pedestrian is situated approximately 26 meters away from the vehicle. Despite this relatively safe distance, the AV initiates a strong braking response, corresponding to the highest deceleration observed throughout the entire trajectory. This contrasts with the behavior in Case 1, where the distance from the pedestrian is comparable, yet the AV adopts a more moderate deceleration strategy. The underlying cause of this difference lies in the initial speed: in this case, the AV starts at a higher velocity, thereby requiring a more forceful reduction in speed to ensure safety. In particular, the peak deceleration reaches -1.458 m/s^2 in Case 3, whereas in Case 1 it only drops to -0.826 m/s^2. Following the initial braking, the AV maintains a conservative speed of around 2 m/s for several seconds, allowing it to continuously reassess its surroundings and avoid potential conflicts. Once pedestrians 1 and 3 have moved sufficiently away from its intended path and pedestrian 2 has already passed, the AV detects a lower risk level and smoothly accelerates to proceed through the environment.

A similar pattern is observed in the behavior of the human driver under the same scenario. The human-operated vehicle also performs an early and substantial deceleration, followed by a sustained period of cautious driving until the nearby pedestrian influence diminishes. As shown in Fig. 5.6c and Fig. 5.6d, the velocity profiles of both the AV and the human-driven vehicle exhibit consistent trends: a sharp initial drop to approximately 2 m/s, followed by a gradual acceleration once the interaction zone is deemed safe. This behavioral consistency highlights the AV's ability to learn and reproduce human-like decision patterns in multi-agent settings.

In conclusion, the above results demonstrate that our proposed AV decision-making model can effectively select appropriate strategies based on the current scenario and state, adjusting its behavior accordingly to achieve safe yet efficient navigation during interactions with pedestrians.

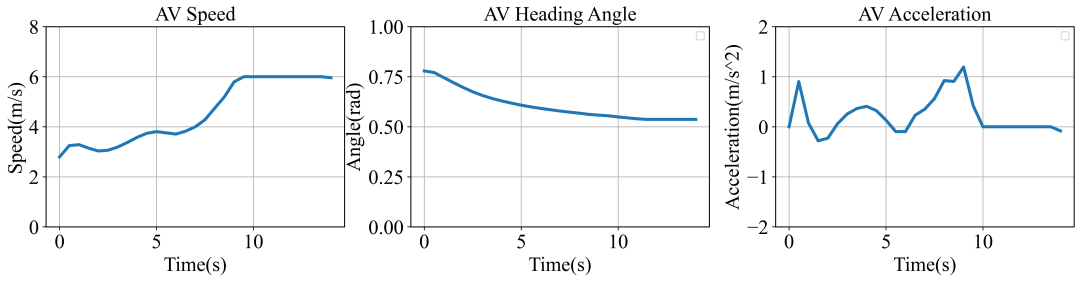


(a)

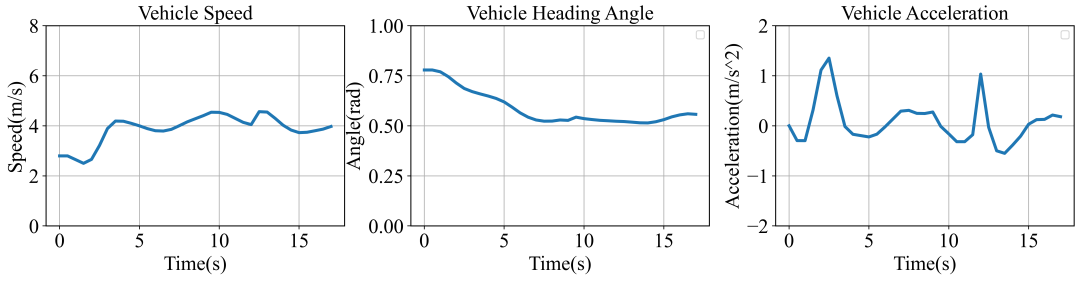


(b)

Figure 5.5: Comparison of the interaction process and dynamic state evolution between the proposed model and real driving data for Case 2. (a) Interaction simulation using the proposed model. (1) State at t = 0 s: v_{AV} = 2.796 m/s, θ = 0.778 rad, a_{AV} = 0 m/s^2 (2) State at t = 6.5 s: v_{AV} = 3.708 m/s, θ = 0.591 rad, a_{AV} = -0.097 m/s^2 (3) State at t = 11 s: v_{AV} = 6.0 m/s, θ = 0.544 rad, a_{AV} = 0 m/s^2 (4) State at t = 15 s: v_{AV} = 5.958 m/s, θ = 0.536 rad, a_{AV} = -0.083 m/s^2. (b) Interaction process captured from real driving data. (1) State at t = 0 s: v_{veh} = 2.796 m/s, θ = 0.778 rad, a_{veh} = 0 m/s^2. (2) State at t = 6.5 s: v_{veh} = 3.802 m/s, θ = 0.564 rad, a_{veh} = -0.024 m/s^2. (3) State at t = 11 s: v_{veh} = 4.454 m/s, θ = 0.531 rad, a_{veh} = -0.317 m/s^2. (4) State at t = 17.5 s: v_{veh} = 3.972 m/s, θ = 0.557 rad, a_{veh} = 0.181 m/s^2.



(c)



(d)

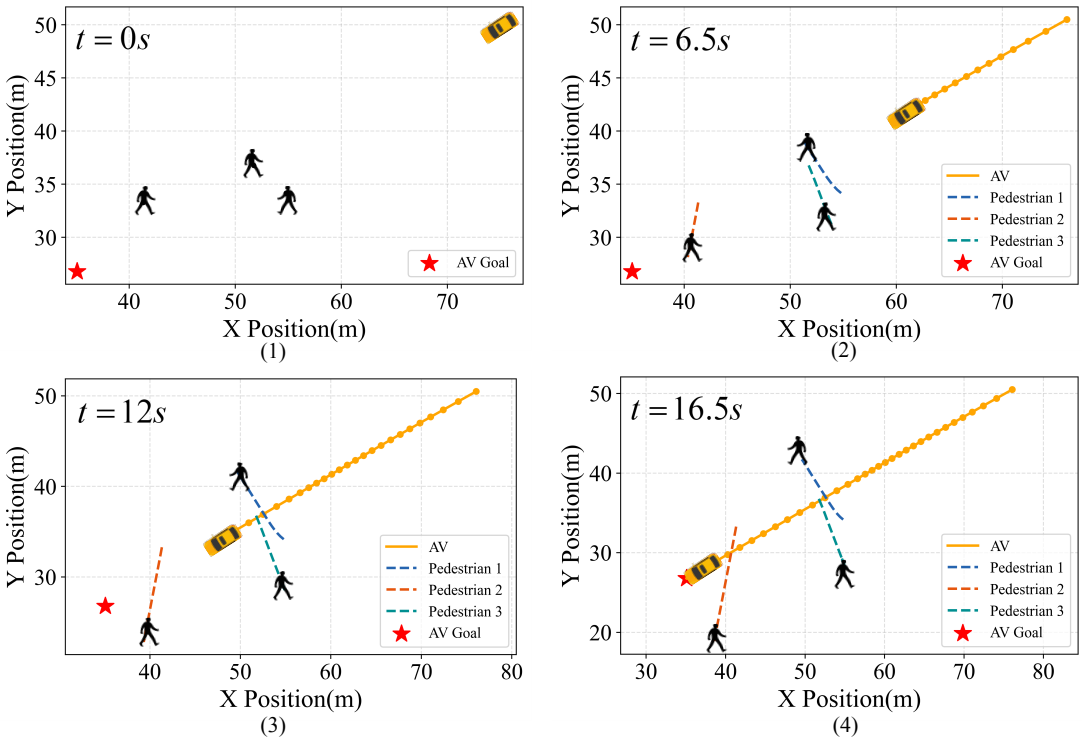
Figure 5.5: (Continued) Comparison of the interaction process and dynamic state evolution between the proposed model and real driving data for Case 2. (c) Evolution of AV states predicted by the proposed model. (d) Evolution of vehicle states recorded from real driving data.

Table 5.3: Quantitative comparison results of our proposed AV model against other approaches and human driving data.

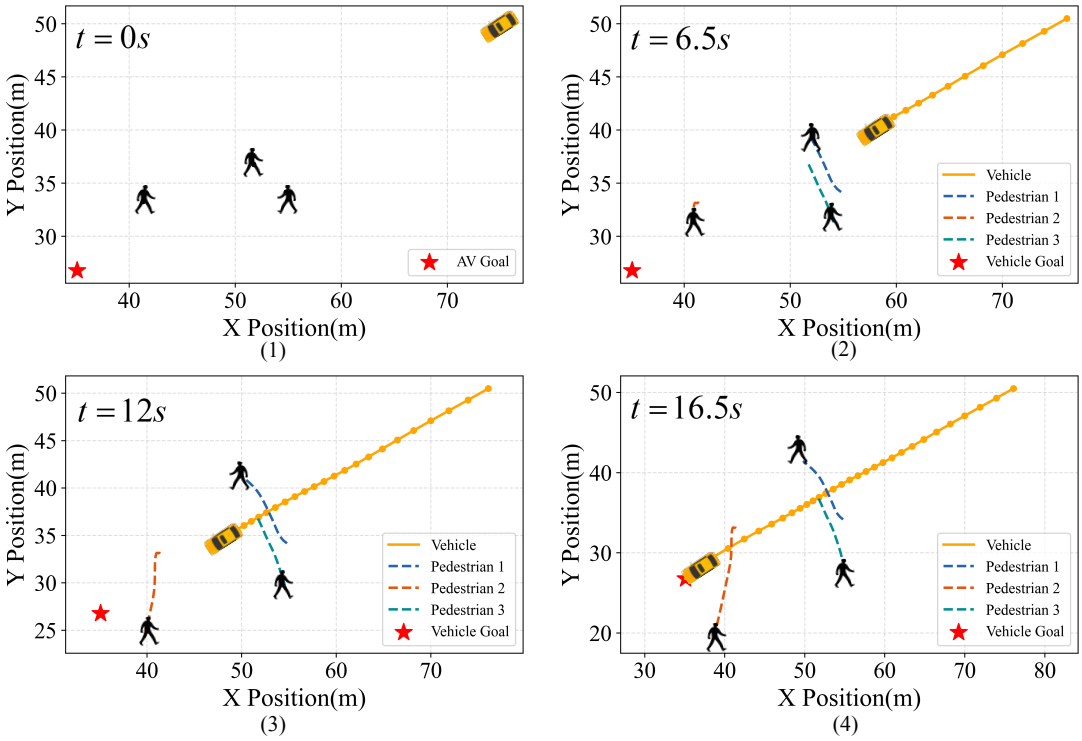
Model	Success	Collision	Time out	Average vehicle speed (m/s)	Average vehicle jerk (m/s^3)	Average maximum absolute acceleration/deceleration (m/s^2)
UAW-PCG [13]	0.83	0.17	0	4.137	1.249	3.737
G-SAC-NoCog	0.83	0.11	0.06	4.035	0.544	1.664
S-SAC	0.78	0.22	0	4.312	0.448	1.560
G-SAC-Cog(Ours)	**0.94**	**0.06**	**0**	**4.470**	**0.524**	**1.621**
Human driver	1	0	0	2.924	0.609	4.240

5.5.2.2 Quantitative Analysis

Table 5.3 summarizes the quantitative performance of our proposed model (G-SAC-Cog) compared to the state-of-the-art method (UAW-PCG), two ablation variants (S-SAC and G-SAC-NoCog) and human driving data. Overall, our method achieves a success rate of 0.94, which is significantly higher than the comparative models (0.78–0.83). Moreover, it attains the lowest collision rate among the AV models at 0.06, indicating that our approach effectively enhances safety in pedestrian-rich environments.

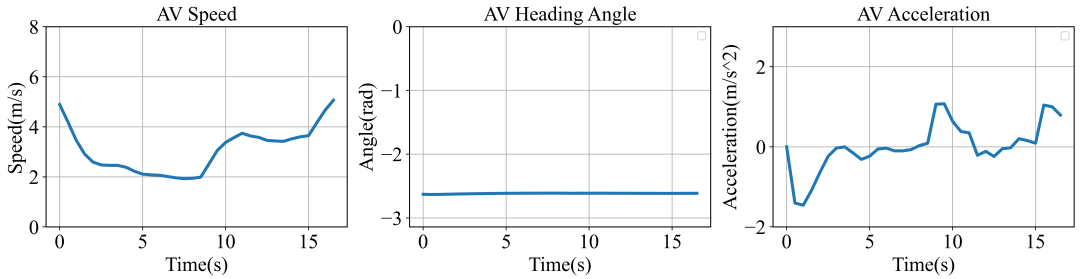


(a)

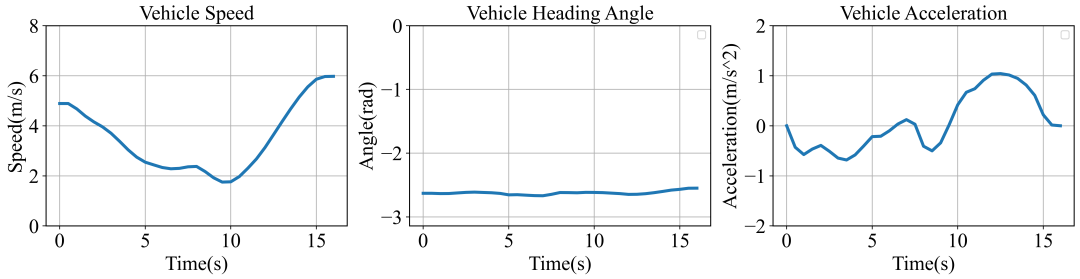


(b)

Figure 5.6: Comparison of the interaction process and dynamic state evolution between the proposed model and real driving data for Case 3. (a) Interaction simulation using the proposed model. (1) State at t = 0 s: v_{AV} = 4.889 m/s, θ = -2.629 rad, a_{AV} = 0 m/s^2 (2) State at t = 6.5 s: v_{AV} = 2.063 m/s, θ = -2.613 rad, a_{AV} = -0.033 m/s^2 (3) State at t = 12 s: v_{AV} = 3.634 m/s, θ = -2.613 rad, a_{AV} = -0.212 m/s^2 (4) State at t = 16.5 s: v_{AV} = 4.664 m/s, θ = -2.614 rad, a_{AV} = 0.999 m/s^2. (b) Interaction process captured from real driving data. (1) State at t = 0 s: v_{veh} = 4.889 m/s, θ = -2.629 rad, a_{veh} = 0 m/s^2. (2) State at t = 6.5 s: v_{veh} = 2.332 m/s, θ = -2.659 rad, a_{veh} = -0.101 m/s^2. (3) State at t = 12 s: v_{veh} = 2.676 m/s, θ = -2.634 rad, a_{veh} = 0.909 m/s^2. (4) State at t = 16.5 s: v_{veh} = 5.975 m/s, θ = -2.549 rad, a_{veh} = 0.0 m/s^2.



(c)



(d)

Figure 5.6: (Continued) Comparison of the interaction process and dynamic state evolution between the proposed model and real driving data for Case 3. (c) Evolution of AV states predicted by the proposed model. (d) Evolution of vehicle states recorded from real driving data.

In terms of driving efficiency, our model achieves an average speed of 4.470 m/s, outperforming all comparative models and greatly exceeding that of the human driver (2.924 m/s). This superior efficiency can be attributed to the reinforcement learning framework applied in training AV models, which, under the given reward structure, incentivizes learned policies to maximize accumulated reward. Consequently, the AV maintains high speeds whenever no immediate collision risk is detected, enabling it to achieve higher efficiency without compromising safety, as evidenced by its low collision rate.

Regarding driving comfort, S-SAC attains the smoothest performance with the minimal average vehicle jerk (0.448 m/s^3). However, this comes with a high collision rate (0.22), suggesting that S-SAC tends to maintain speed even in close proximity to pedestrians, resulting in smoother but riskier behavior. Conversely, UAW-PCG shows the most erratic driving style, with an average vehicle jerk of 1.249 m/s^3. This instability likely stems from directly controlling vehicle speed, which forces the vehicle to make sudden and large acceleration or deceleration adjustments to match target speeds, resulting in frequent and abrupt speed variations. Our proposed model, which employs acceleration as the control action, achieves a well-balanced performance. Although its average jerk (0.524 m/s^3) is not the lowest, it reflects necessary adjustments for effective pedestrian avoidance and remains lower than that of the human driver (0.609 m/s^3), successfully balancing safety and comfort. When considering maximum absolute acceleration or deceleration, all

AV models display substantially lower peaks compared to the human driver (4.240 m/s^2), indicating smoother control policies acquired through reinforcement learning. Notably, UAW-PCG shows a significantly elevated average maximum acceleration (3.737 m/s^2), again reflecting instability from using velocity as the direct action variable.

In conclusion, these findings demonstrate that our proposed model achieves a superior balance between efficiency, safety, and driving comfort compared to existing methods.

5.6 Chapter Conclusion

This chapter begins by identifying two key limitations in current research on AV decision-making in urban shared environments. First, many existing approaches to AV-pedestrian interactions employ overly simplified pedestrian models with fixed behavioral patterns, which fail to capture the inherently bidirectional and adaptive nature of human interactions. Second, although robot navigation in crowd settings has been extensively studied, the subjective uncertainty underlying human cognition has received insufficient attention. To address these gaps, this study presents an innovative framework to address AV decision-making challenges in shared urban spaces with multiple pedestrians. A cognitive uncertainty modeling approach inspired by the Free Energy Principle is introduced to emulate human reasoning and quantify cognitive uncertainty during social interactions. This cognitive modeling is integrated into the pedestrian social force model, enabling adaptive adjustment of goal-directed and repulsive forces by combining cognitive uncertainty with physical risk, which leads to more realistic pedestrian trajectories. Additionally, the same cognitive process guides the AV's decision-making, where the fused measure of cognitive uncertainty and physical risk defines a dynamic, risk-aware adjacency matrix for the GCN within the SAC architecture, allowing the AV to perceive complex social dynamics and make human-like decisions.

Simulation results demonstrate that the proposed framework significantly improves AV performance compared to the existing method, achieving safer, more efficient, and smoother behavior during interactions with pedestrians. Although the model shows promising results, the current research is limited to scenarios with a fixed number of pedestrians, which may not fully capture the variability of real-world urban environments. In the future, we plan to extend our framework to handle scenarios with dynamically changing numbers of pedestrians, enabling the AV to adapt its strategy in real time as pedestrians enter or leave the scene. Moreover, this work focuses on a simplified interaction scenario, so certain real-world factors, such as road structures, weather conditions, and the presence of additional vulnerable road users, are not considered. Future work should incorporate

these elements to improve the generalizability and applicability of the proposed models. Although the RL demonstrates good performance in this work, it still faces several limitations. RL algorithms are highly sensitive to the design of the reward function, and inappropriate shaping can lead to undesired behaviors. In addition, training may suffer from instability due to exploration-exploitation trade-offs and approximation errors. Future work should investigate more principled reward-design strategies and stabilization techniques tailored to the specific task.

Chapter 6

Conclusion

6.1 Summary of Research

This thesis is dedicated solely to AV-pedestrian interactions, spanning scenarios from simple single-pedestrian to complex multi-pedestrian situations. The primary objective is to gain a deeper understanding of the interaction mechanisms and to develop decision-making frameworks that ensure navigation strategies which are safe, efficient, and socially acceptable in human-centric environments. The study is structured around three main components: interaction data analysis, AV decision-making in single-pedestrian interactions, and AV decision-making in multi-pedestrian interactions.

Interaction Data Analysis. This study investigates the critical factors influencing the decision-making processes of both human drivers and pedestrians. A series of vehicle–pedestrian interaction experiments using the VR platform were conducted, enabling the acquisition of controlled yet realistic behavioral data. The analysis primarily focused on kinematic variables, as they are easily measurable during real-world interactions. Using the collected data, the AdaBoost algorithm was applied to identify the most influential factors affecting pedestrian crossing decisions and driver approaching behaviors. Furthermore, PDPs were utilized to visualize and interpret the impact of each factor, illustrating how variations in these variables affect decision-making outcomes. The analysis revealed that pedestrian crossing intentions are mainly influenced by longitudinal distance and vehicle acceleration, while driver yielding behavior is strongly affected by pedestrian speed and their longitudinal distance to the vehicle. Specifically, longer longitudinal distances encourage pedestrians to cross before vehicles approach, with crossing probability increasing as distance grows. By contrast, higher vehicle acceleration reduces the likelihood of pedestrians crossing, and pedestrians are particularly more likely to cross when the vehicle's acceleration is below –1 m/s^2. For drivers, the probability of yielding increases with higher pedestrian walking speed and also rises with longer longitudinal distances. In addition, this work investigates drivers' yielding patterns. The results indicated that when

a pedestrian's intention to cross is ambiguous, most drivers slightly reduce their speed, even under low TTA conditions and more than half of the drivers do not decelerate at all under high TTA scenarios. Conversely, when a pedestrian's intention is clearly expressed, a significant proportion of drivers adopt moderate or heavy deceleration, particularly in low TTA situations. For TTA values greater than 5 seconds, light deceleration becomes the dominant behavior. These findings emphasized the critical role of kinematic cues in shaping both pedestrian and driver decision-making during interactions.

AV decision-making in single-pedestrian interactions. This study examines the decision-making strategies of the AV when interacting with a single pedestrian in unsignalized intersections, with a particular focus on incorporating human bounded rationality. A dynamic, real-time decision-making framework was proposed for AV navigation in unsignalized environments. The proposed framework integrated POMDP with behavioral game theory to capture continuous and interactive decision-making between AVs and pedestrians. Both the AV and pedestrian behaviors were represented using a DB-QCH model, which consists of three core modules: belief update, action generation, and optimal action derivation. To better simulate human decision-making, two cognitive parameters, the reasoning level k and the bounded rationality λ, are incorporated and dynamically updated based on the observed environment. This adaptive mechanism enables the AV to continuously refine its strategies and respond appropriately in real-time.

The proposed model was validated through both qualitative and quantitative analyses. Qualitative evaluations include three representative scenarios illustrating how the AV adapts its behavior under varying pedestrian interactions. Visualizations of AV and pedestrian states, such as position and velocity profiles, showed the AV's ability to anticipate pedestrian intentions and apply appropriate deceleration, demonstrating robust decision-making across diverse interaction contexts. Quantitative evaluations compared the model's performance against data from VR experiments, focusing on safety, efficiency, and smoothness. The results confirmed that the algorithm performs strongly across all three metrics. Specifically, the model achieved a very low collision rate of 0.15%. The close alignment between simulated and observed vehicle speeds indicated that the model captures key features of human driving behavior, while its average speed reached 9.60 m/s, slightly higher than the 9.47 m/s observed in VR experiments, reflecting improved efficiency. Furthermore, compared with the VR experiment, the average vehicle jerk decreased by 0.054 m/s^3, and the average maximum absolute acceleration/deceleration was reduced by 0.361 m/s^2. The reductions demonstrated the strategy's ability to generate smoother, more comfortable driving trajectories than those observed in human-operated data.

AV decision-making in multi-pedestrian interactions. This study investigates the decision-making strategies of the AV operating in pedestrian-rich shared spaces, with a particular emphasis on modeling cognitive uncertainty. A cognitively-informed and risk-aware decision-making framework for the AV navigating dynamic pedestrian environments was presented. Inspired by the Free Energy Principle, a cognitive uncertainty modeling approach was proposed to simulate human reasoning processes and quantify uncertainty in agents' predictions. By computing the KL divergence between predicted and observed state distributions, the model enables agents to iteratively update their beliefs about others' behaviors and gain a better understanding of their surroundings. Building on this foundation, a novel pedestrian behavior model, called Cognitive-Risk Social Force Model was introduced. This model dynamically adjusts the weights of goal-directed and repulsive forces based on a fused measure of physical risk and cognitive uncertainty, allowing pedestrians to autonomously adapt their trajectories in uncertain and interactive scenarios and resulting in more realistic, human-like behavior. To support AV decision-making, the fused risk signal was further integrated into a graph-based reinforcement learning architecture. A risk-encoded interaction graph was constructed and embedded into both the actor and critic networks of a SAC agent. This structure enables the AV to more effectively perceive interaction dynamics and respond to potential threats, thereby improving decision quality.

The effectiveness of the proposed pedestrian CR-SFM and AV decision-making model was thoroughly evaluated through a series of qualitative and quantitative experiments. For the pedestrian CR-SFM, three representative case studies were conducted to visualize and compare its simulated trajectories with those generated by the traditional SFM and its variant (RA-SFM). These qualitative comparisons demonstrated that CR-SFM produces more adaptive and reliable motion patterns under diverse and challenging conditions. To further assess its performance, a quantitative evaluation was carried out using three widely adopted metrics: ADE, FDE, and CR. Across all metrics, CR-SFM consistently outperformed the baselines, achieving lower positional errors and fewer collisions. In particularly, relative to the CV model, ADE and FDE were reduced by 48.6% and 84.6%, respectively. Compared with both the SFM and RA-SFM, CR-SFM further lowered ADE (by up to 5.6%) and FDE (by up to 28.2%). For all comparison models, the collision rate of CR-SFM dropped to zero. These results highlighted its improved capability to accurately capture pedestrian behaviors and adapt effectively to dynamic, multi-agent environments. The AV decision-making model was also validated through three case-based comparisons with real-world human driving data. Each case examined the AV's behavior in terms of trajectory, velocity, heading angle, and acceleration profile, capturing a range of driving scenarios involving varying degrees of risk and interaction complexity. The qualitative analysis revealed that the AV demonstrated human-like decision-making, including proactive yielding, smooth acceleration, and adaptive braking in response to pedestrian movement.

Beyond these illustrative examples, a quantitative evaluation was conducted by validate the proposed AV model against a state-of-the-art baseline and two ablated versions. The assessment employed three key performance dimensions, including safety, efficiency, and comfort, to capture the overall driving quality. Results showed that the proposed framework achieved lowest collision rates, highest average speeds, and reduced vehicle jerk. In particular, compared with the state-of-the-art model, the proposed approach achieved a 13.2% increase in success rate and an 8% improvement in average speed, while the average jerk decreased by 58% and the maximum absolute acceleration/deceleration by 56.6%. Moreover, relative to human drivers, it demonstrated greater efficiency through higher average speed and delivered smoother driving with lower jerk and reduced maximum acceleration/deceleration. These outcomes confirm that the model can dynamically select context-appropriate strategies and make adjustments to its behavior in response to environmental changes.

In summary, this study contributes to the field of AV-pedestrian interaction by identifying the key factors that influence the decision-making processes of both human drivers and pedestrians, developing an interactive model for unsignalized intersections involving an AV and a single pedestrian, and extending this framework to urban shared spaces with multiple pedestrians to capture more complex interaction dynamics. These contributions effectively address the three research questions outlined in this thesis and provide valuable insights as well as practical solutions to the challenges of AV decision-making in human-involved urban environments.

6.2 Limitations and Future Work

While this work has made notable progress in modeling AV–pedestrian interactions and developing decision-making frameworks, several limitations remain due to the assumptions and scope of the current study. These limitations provide opportunities for future research aimed at further improving the overall performance of the proposed models. The key limitations and potential directions for future work are summarized as follows:

Integrating non-verbal cues in AV decision-making: Although the proposed AV models can make high-level decisions, such as determining whether to cross or yield to pedestrians, they currently rely exclusively on fundamental kinematic factors, including pedestrian positions and velocities. These parameters form a solid foundation for modeling interactions. However, in real-world scenarios, pedestrian intentions are often conveyed not only through motion patterns but also through subtle non-verbal signals, such as gaze direction, head orientation, hand gestures, and subtle body movements, which often

provide additional context regarding a pedestrian's willingness or hesitation to cross. For instance, direct eye contact or a forward-leaning posture toward an approaching vehicle often indicates a strong willingness to cross, while avoiding eye contact or stepping back typically signals hesitation or a decision to yield.

Future work could focus on augmenting the AV's decision-making model by integrating such non-verbal cues with existing kinematic information, enabling the AV to form a more comprehensive understanding of pedestrian behavior. By combining these complementary data sources, the AV could better anticipate pedestrian actions and react more appropriately, particularly in ambiguous or high-risk interaction scenarios. Furthermore, leveraging advanced perception techniques, such as gaze tracking or gesture recognition, could enhance both the interpretability and social awareness of AV decision-making, thereby improving trust, safety, and overall performance in complex urban environments.

Scalability to dynamic and group pedestrian interactions: The AV–pedestrian interaction framework proposed in this thesis is currently restricted to scenarios involving an AV and three pedestrians, primarily due to computational limitations on local hardware. In real-world urban settings, however, the number of pedestrians can easily exceed three and often varies dynamically. For example, five pedestrians may be observed at one moment, increasing to eight or more in the next. Such fluctuations introduce additional complexity in decision-making, which the current model does not fully capture. Moreover, our existing approach models pedestrians as independent agents, while in reality, pedestrians often move in groups. Intra-group dynamics, such as leader–follower behaviors, coordinated movements, and collective decision-making can significantly influence the overall interaction patterns. These group behaviors can affect factors like crossing decisions, yielding patterns, and perceived risks, making individual-based modeling insufficient in some situations.

Future work could focus on extending the current framework to accommodate dynamically varying pedestrian populations and group-level behaviors. This may involve developing scalable graph-based interaction models that can dynamically adapt to changing pedestrian populations, as well as introducing group features that capture shared intentions and coordinated movement patterns. Such extensions would enable AVs to operate safely and effectively in more realistic, crowded, and socially complex environments.

Interaction modeling with diverse road users: This thesis focuses primarily on AV–pedestrian interactions, while in real-world traffic scenarios, AV must account for a broader range of road users. In addition to pedestrians, human-driven vehicles and other vulnerable road users, such as cyclists, motorcyclists, wheelchair users, frequently share the same road space with AVs, introducing additional complexity to decision-making.

These road users differ not only in their kinematic characteristics, such as moving speed, acceleration, and braking distance, but also in their behavioral patterns and responses to AV actions. For instance, cyclists may perform sudden lateral lane changes to avoid obstacles or maintain balance, human-driven vehicles may exhibit unpredictable behaviors such as abrupt acceleration or delayed braking, and wheelchair users generally move at lower speeds with limited turning radii, requiring the AV to adopt more conservative predictions and maintain larger safety buffers.

Future work could extend the current decision-making framework to explicitly address interactions with multiple types of road users. This may involve developing multi-agent models that capture heterogeneous dynamics and diverse behavioral patterns, thereby improving the AV's ability to make decisions and adapt to complex and uncertain interactions. Such modeling would enhance the robustness and social compliance of AV decision-making in mixed-traffic urban environments and accelerate the deployment of AV technologies in real-world scenarios.

Generalization ability across various environments: This thesis mainly investigates interactions between the AV and pedestrians in specific environments, such as unsignalized intersections and urban shared spaces. While the proposed models perform effectively in these settings, they are primarily tailored to address individual scenario types rather than being adaptable across diverse traffic conditions. However, real-world driving involves navigating a wide variety of environments, ranging from highways and rural roads to densely populated city centers. During a single driving journey, an AV may encounter multiple environment types, each with unique dynamics and challenges. Seamlessly adapting to these diverse and evolving scenarios remains a significant challenge for the current framework.

Future work could focus on enhancing the adaptability and generalization capability of AV decision-making models to handle diverse real-world conditions. This may involve developing a more versatile decision-making framework that dynamically adjusts its strategies in response to contextual factors such as traffic density, road geometry, and the presence of various road users. Achieving this generalization would represent a critical step toward deploying AVs that can ensure safe and efficient navigation across all types of real-world traffic environments.

Real-world experiment validation: The AV decision-making algorithms developed in this thesis have been thoroughly evaluated and validated within a self-constructed simulation environment. While these models have demonstrated strong performance across a range of qualitative and quantitative metrics, simulation-based evaluations inevitably have limitations in replicating the full complexity of real-world conditions. Elements such

as sensor perception errors, natural variability in human behavior, and environmental factors (e.g., weather or road conditions) are often simplified during simulation. For example, our experiments assume complete and accurate access to pedestrian positions and velocities, whereas real-world sensors may introduce subtle inaccuracies, and pedestrian behaviors may deviate from those represented in the interaction models. Therefore, conducting real-world experiments would be a valuable complement, offering further evidence of the proposed models' reliability and adaptability under actual operational conditions. Future work could focus on integrating these decision-making models into real AV platforms and performing field tests in dynamic urban environments. This would provide a more comprehensive evaluation of the models' robustness, safety, and ability to handle dynamic and complex real-world scenarios.

Bibliography

[1] M. Sadaf et al., 'Connected and automated vehicles: Infrastructure, applications, security, critical challenges, and future aspects', *Technologies*, vol. 11, no. 5, p. 117, 2023.

[2] S. O.-R. A. V. S. Committee et al., 'Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems', *SAE Standard J*, vol. 3016, p. 1, 2014.

[3] Q. Liu, X. Li, S. Yuan and Z. Li, 'Decision-making technology for autonomous vehicles: Learning-based methods, applications and future outlook', in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 30–37.

[4] M. M. Rahman and J.-C. Thill, 'Impacts of connected and autonomous vehicles on urban transportation and environment: A comprehensive review', *Sustainable Cities and Society*, vol. 96, p. 104 649, 2023.

[5] R. A. Khalil, Z. Safelnasr, N. Yemane, M. Kedir, A. Shafiqurrahman and N. Saeed, 'Advanced learning technologies for intelligent transportation systems: Prospects and challenges', *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 397–427, 2024.

[6] X. Li, S.-A. Kaye, A. P. Afghari and O. Oviedo-Trespalacios, 'Sharing roads with automated vehicles: A questionnaire investigation from drivers', cyclists' and pedestrians' perspectives', *Accident Analysis & Prevention*, vol. 188, p. 107 093, 2023.

[7] M. T. Rahman, K. Dey, S. Das and M. Sherfinski, 'Sharing the road with autonomous vehicles: A qualitative analysis of the perceptions of pedestrians and bicyclists', *Transportation research part F: traffic psychology and behaviour*, vol. 78, pp. 433–445, 2021.

[8] A. Mirzabagheri, M. Ahmadi, N. Zhang, R. Alirezaee, S. Mozaffari and S. Alirezaee, 'Navigating uncertainty: Advanced techniques in pedestrian intention prediction for autonomous vehicles—a comprehensive review', *Vehicles*, vol. 7, no. 2, p. 57, 2025.

[9] Y. Liu, Y. Lyu, K. Böttcher and M. Rötting, 'External interface-based autonomous vehicle-to-pedestrian communication in urban traffic: Communication needs and design considerations', *International Journal of Human–Computer Interaction*, vol. 36, no. 13, pp. 1258–1272, 2020.

[10] F Pascucci, N Rinke, C Schiermeyer, V Berkhahn and B Friedrich, 'A discrete choice model for solving conflict situations between pedestrians and vehicles in shared space', *arXiv preprint arXiv:1709.09412*, 2017.

[11] K. Tian et al., 'Deceleration parameters as implicit communication signals for pedestrians' crossing decisions and estimations of automated vehicle behaviour', *Accident Analysis & Prevention*, vol. 190, p. 107 173, 2023.

[12] C. Schöller, V. Aravantinos, F. Lay and A. Knoll, 'What the constant velocity model can teach us about pedestrian motion prediction', *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1696–1703, 2020.

[13] M. Golchoubian, M. Ghafurian, K. Dautenhahn and N. L. Azad, 'Uncertainty-aware drl for autonomous vehicle crowd navigation in shared space', *IEEE Transactions on Intelligent Vehicles*, 2024.

[14] Y. Chen, S. Li, X. Tang, K. Yang, D. Cao and X. Lin, 'Interaction-aware decision-making for autonomous vehicles', *IEEE Transactions on Transportation Electrification*, vol. 9, no. 3, pp. 4704–4715, 2023.

[15] G. P. R. Papini, A. Plebe, M. Da Lio and R. Donà, 'A reinforcement learning approach for enacting cautious behaviours in autonomous driving system: Safe speed choice in the interaction with distracted pedestrians', *IEEE transactions on intelligent transportation systems*, vol. 23, no. 7, pp. 8805–8822, 2021.

[16] F. Hafeez, U. Ullah Sheikh, A. A. Mas' ud, S. Al-Shammari, M. Hamid and A. Azhar, 'Application of the theory of planned behavior in autonomous vehicle-pedestrian interaction', *Applied Sciences*, vol. 12, no. 5, p. 2574, 2022.

[17] A. Esmaili, K. Aghabayk, N. Parishad and A. N. Stephens, 'Investigating the interaction between pedestrian behaviors and crashes through validation of a pedestrian behavior questionnaire (pbq)', *Accident Analysis & Prevention*, vol. 153, p. 106 050, 2021.

[18] L. Govinda, M. S. K. Raju and K. R. Shankar, 'Pedestrian-vehicle interaction severity level assessment at uncontrolled intersections using machine learning algorithms', *Safety science*, vol. 153, p. 105 806, 2022.

[19] W. Chen, T. Wang, Y. Wang, Q. Li, Y. Xu and Y. Niu, 'Lane-based distance-velocity model for evaluating pedestrian-vehicle interaction at non-signalized locations', *Accident Analysis & Prevention*, vol. 176, p. 106 810, 2022.

[20] V. Gagliardi, C. Ferrante and F. Bella, 'Safety assessment of pedestrian-vehicle interaction at signalized intersections: An observational study', *Journal of Transportation Safety & Security*, vol. 16, no. 10, pp. 1212–1235, 2024.

[21] A. Rasch, G. Panero, C.-N. Boda and M. Dozza, 'How do drivers overtake pedestrians? evidence from field test and naturalistic driving data', *Accident Analysis & Prevention*, vol. 139, p. 105 494, 2020.

[22] S. Sun, Z. Zhang, Z. Zhang, P. Deng, K. Tian and C. Wei, 'How do human-driven vehicles avoid pedestrians in interactive environments? a naturalistic driving study', *Sensors*, vol. 22, no. 20, p. 7860, 2022.

[23] T. Z. Noonan, P. Gershon, J. Domeyer, B. Mehler and B. Reimer, 'Kinematic cues in driver-pedestrian communication to support safe road crossing', *Accident Analysis & Prevention*, vol. 192, p. 107 236, 2023.

[24] T. T. M. Tran, C. Parker and M. Tomitsch, 'A review of virtual reality studies on autonomous vehicle–pedestrian interaction', *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 6, pp. 641–652, 2021.

[25] D. T. Luu, H. Eom, G.-H. Cho, S.-N. Kim, J. Oh and J. Kim, 'Cautious behaviors of pedestrians while crossing narrow streets: Exploration of behaviors using virtual reality experiments', *Transportation research part F: traffic psychology and behaviour*, vol. 91, pp. 164–178, 2022.

[26] F. Camara, P. Dickinson and C. Fox, 'Evaluating pedestrian interaction preferences with a game theoretic autonomous vehicle in virtual reality', *Transportation research part F: traffic psychology and behaviour*, vol. 78, pp. 410–423, 2021.

[27] L. Crosato, C. Wei, E. S. Ho, H. P. Shum and Y. Sun, 'A virtual reality framework for human-driver interaction research: Safe and cost-effective data collection', in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 167–174.

[28] Y. Feng, Z. Xu, H. Farah and B. Van Arem, 'Does another pedestrian matter? a virtual reality study on the interaction between multiple pedestrians and autonomous vehicles in shared space', *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[29] A. Baldassa, F. Orsini, G. De Cet, M. Tagliabue, R. Rossi and M. Gastaldi, 'Validation of an urban environment for pedestrian behavior analysis in full immersive virtual reality', *Transportation research procedia*, vol. 78, pp. 24–31, 2024.

[30] B. Nie, Q. Li, S. Gan, B. Xing, Y. Huang and S. E. Li, 'Safety envelope of pedestrians upon motor vehicle conflicts identified via active avoidance behaviour', *Scientific reports*, vol. 11, no. 1, p. 3996, 2021.

[31] C. Holland and R. Hill, 'The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations', *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 224–237, 2007.

[32] K. Wilmut and C. Purcell, 'Why are older adults more at risk as pedestrians? a systematic review', *Human factors*, vol. 64, no. 8, pp. 1269–1291, 2022.

[33] A. H. Kalantari et al., 'Who goes first? a distributed simulator study of vehicle–pedestrian interaction', *Accident Analysis & Prevention*, vol. 186, p. 107 050, 2023.

[34] H. Wang, A. Wang, F. Su and D. C. Schwebel, 'The effect of age and sensation seeking on pedestrian crossing safety in a virtual reality street', *Transportation research part F: traffic psychology and behaviour*, vol. 88, pp. 99–110, 2022.

[35] I. E. Hyman Jr, S. M. Boss, B. M. Wise, K. E. McKenzie and J. M. Caggiano, 'Did you see the unicycling clown? inattentional blindness while walking and talking on a cell phone', *Applied Cognitive Psychology*, vol. 24, no. 5, pp. 597–607, 2010.

[36] E. Y. Du et al., 'Pedestrian behavior analysis using 110-car naturalistic driving data in usa', in *23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, vol. 13, 2013, pp. 27–30.

[37] C. M. DiPietro and L. E. King, 'Pedestrian gap-acceptance', *Highway Research Record*, no. 308, 1970.

[38] S. Schmidt and B. Faerber, 'Pedestrians at the kerb–recognising the action intentions of humans', *Transportation research part F: traffic psychology and behaviour*, vol. 12, no. 4, pp. 300–310, 2009.

[39] X. Xin, N. Jia, S. Ling and Z. He, 'Prediction of pedestrians' wait-or-go decision using trajectory data based on gradient boosting decision tree', *Transportmetrica B: transport dynamics*, vol. 10, no. 1, pp. 693–717, 2022.

[40] A. Sheykhfard and F. Haghighi, 'Performance analysis of urban drivers encountering pedestrian', *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 62, pp. 160–174, 2019.

[41] K. Tian et al., 'Explaining unsafe pedestrian road crossing behaviours using a psychophysics-based gap acceptance model', *Safety science*, vol. 154, p. 105 837, 2022.

[42] J. Zhao, J. O. Malenje, Y. Tang and Y. Han, 'Gap acceptance probability model for pedestrians at unsignalized mid-block crosswalks based on logistic regression', *Accident Analysis & Prevention*, vol. 129, pp. 76–83, 2019.

[43] M. M. Ishaque and R. B. Noland, 'Behavioural issues in pedestrian speed choice and street crossing behaviour: A review', *Transport Reviews*, vol. 28, no. 1, pp. 61–85, 2008.

[44] J. Zhao, J. O. Malenje, J. Wu and R. Ma, 'Modeling the interaction between vehicle yielding and pedestrian crossing behavior at unsignalized midblock crosswalks', *Transportation research part F: traffic psychology and behaviour*, vol. 73, pp. 222–235, 2020.

[45] M. M. Hamed, 'Analysis of pedestrians' behavior at pedestrian crossings', *Safety science*, vol. 38, no. 1, pp. 63–82, 2001.

[46] B. R. Kadali and P. Vedagiri, 'Proactive pedestrian safety evaluation at unprotected mid-block crosswalk locations under mixed traffic conditions', *Safety science*, vol. 89, pp. 94–105, 2016.

[47] M. Al Eisaeia, S. Moridpourb and R. Tay, 'Heavy vehicle management: Restriction strategies', *Transportation Research Procedia*, vol. 21, pp. 18–28, 2017.

[48] S. Das, C. F. Manski and M. D. Manuszak, 'Walk or wait? an empirical analysis of street crossing decisions', *Journal of applied econometrics*, vol. 20, no. 4, pp. 529–548, 2005.

[49] A. Rasouli, I. Kotseruba and J. K. Tsotsos, 'Agreeing to cross: How drivers and pedestrians communicate', in *2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2017, pp. 264–269.

[50] K. Aghabayk, J. Esmailpour, A. Jafari and N. Shiwakoti, 'Observational-based study to explore pedestrian crossing behaviors at signalized and unsignalized crosswalks', *Accident Analysis & Prevention*, vol. 151, p. 105 990, 2021.

[51] R. Sun, X. Zhuang, C. Wu, G. Zhao and K. Zhang, 'The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment', *Transportation research part F: traffic psychology and behaviour*, vol. 30, pp. 97–106, 2015.

[52] J. P. N. Velasco et al., 'Will pedestrians cross the road before an automated vehicle? the effect of drivers' attentiveness and presence on pedestrians' road crossing behavior', *Transportation research interdisciplinary perspectives*, vol. 12, p. 100 466, 2021.

[53] Y. M. Lee et al., 'Road users rarely use explicit communication when interacting in today's traffic: Implications for automated vehicles', *Cognition, Technology & Work*, vol. 23, no. 2, pp. 367–380, 2021.

[54] N. AbuAli and H. Abou-Zeid, 'Driver behavior modeling: Developments and future directions', *International journal of vehicular technology*, vol. 2016, no. 1, p. 6 952 791, 2016.

[55] L. Jing, W. Shan and Y. Zhang, 'Risk preference, risk perception as predictors of risky driving behaviors: The moderating effects of gender, age, and driving experience', *Journal of Transportation Safety & Security*, vol. 15, no. 5, pp. 467–492, 2023.

[56] S. H. Hamdar, L. Qin and A. Talebpour, 'Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework', *Transportation research part C: emerging technologies*, vol. 67, pp. 193–213, 2016.

[57] M. V. Faria, P. C. Baptista, T. L. Farias and J. M. Pereira, 'Assessing the impacts of driving environment on driving behavior patterns', *Transportation*, vol. 47, no. 3, pp. 1311–1337, 2020.

[58] L. Li, J. Gan, Z. Yi, X. Qu and B. Ran, 'Risk perception and the warning strategy based on safety potential field theory', *Accident Analysis & Prevention*, vol. 148, p. 105 805, 2020.

[59] S. Kolekar, J. De Winter and D. Abbink, 'Human-like driving behaviour emerges from a risk-based driver model', *Nature communications*, vol. 11, no. 1, p. 4850, 2020.

[60] X. Zhao, R. He and J. Wang, 'How do drivers respond to driving risk during car-following? risk-response driver model and its application in human-like longitudinal control', *Accident Analysis & Prevention*, vol. 148, p. 105 783, 2020.

[61] J. E. Domeyer, J. D. Lee, H. Toyoda, B. Mehler and B. Reimer, 'Driver-pedestrian perceptual models demonstrate coupling: Implications for vehicle automation', *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 557–566, 2022.

[62] Q. Xue, G. Markkula, X. Yan and N. Merat, 'Using perceptual cues for brake response to a lead vehicle: Comparing threshold and accumulator models of visual looming', *Accident Analysis & Prevention*, vol. 118, pp. 114–124, 2018.

[63] A. Zgonnikov, D. Abbink and G. Markkula, 'Should i stay or should i go? cognitive modeling of left-turn gap acceptance decisions in human drivers', *Human factors*, vol. 66, no. 5, pp. 1399–1413, 2024.

[64] Z. Wang et al., 'Driver behavior modeling using game engine and real vehicle: A learning-based approach', *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 4, pp. 738–749, 2020.

[65] J. Wang, Z. Zhang and G. Lu, 'A bayesian inference based adaptive lane change prediction model', *Transportation research part C: emerging technologies*, vol. 132, p. 103 363, 2021.

[66] J. Ma, H. Xie, K. Song and H. Liu, 'A bayesian driver agent model for autonomous vehicles system based on knowledge-aware and real-time data', *Sensors*, vol. 21, no. 2, p. 331, 2021.

[67] M. Fukui and Y. Ishibashi, 'Self-organized phase transitions in cellular automaton models for pedestrians', *Journal of the physical society of Japan*, vol. 68, no. 8, pp. 2861–2863, 1999.

[68] P. Fiorini and Z. Shiller, 'Motion planning in dynamic environments using velocity obstacles', *The international journal of robotics research*, vol. 17, no. 7, pp. 760–772, 1998.

[69] J. Van Den Berg, S. J. Guy, M. Lin and D. Manocha, 'Reciprocal n-body collision avoidance', in *Robotics Research: The 14th International Symposium ISRR*, Springer, 2011, pp. 3–19.

[70] C. Chen, Y. Liu, S. Kreiss and A. Alahi, 'Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning', in *2019 international conference on robotics and automation (ICRA)*, IEEE, 2019, pp. 6015–6022.

[71] D. Helbing and P. Molnar, 'Social force model for pedestrian dynamics', *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

[72] M. Kabtoul, M. Prédhumeau, A. Spalanzani, J. Dugdale and P. Martinet, 'How to evaluate the navigation of autonomous vehicles around pedestrians?', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 2311–2321, 2023.

[73] M. M. Rashid, M. Seyedi and S. Jung, 'Simulation of pedestrian interaction with autonomous vehicles via social force model', *Simulation modelling practice and theory*, vol. 132, p. 102 901, 2024.

[74] S. Li, W. Li, C. Cook, C. Zhu and Y. Gao, 'Independently recurrent neural network (indrnn): Building a longer and deeper rnn', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5457–5466.

[75] X. Song et al., 'Pedestrian trajectory prediction based on deep convolutional lstm network', *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3285–3302, 2020.

[76] R. Quan, L. Zhu, Y. Wu and Y. Yang, 'Holistic lstm for pedestrian trajectory prediction', *IEEE transactions on image processing*, vol. 30, pp. 3229–3239, 2021.

[77] Y. Zhang and L. Zheng, 'Pedestrian trajectory prediction with mlp-social-gru', in *Proceedings of the 2021 13th International Conference on Machine Learning and Computing*, 2021, pp. 368–372.

[78] Z. Xu, Q. Yu, W. Slamu, Y. Zhou and Z. Liu, 'S-cgru: An efficient model for pedestrian trajectory prediction', in *International Conference on Neural Information Processing*, Springer, 2023, pp. 244–259.

[79] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, 'Social lstm: Human trajectory prediction in crowded spaces', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.

[80] Z. Jiang et al., 'Social nstransformers: Low-quality pedestrian trajectory prediction', *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 11, pp. 5575–5588, 2024.

[81] L. Li, M. Pagnucco and Y. Song, 'Graph-based spatial transformer with memory replay for multi-future pedestrian trajectory prediction', in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 2231–2241.

[82] L. Shi, L. Wang, S. Zhou and G. Hua, 'Trajectory unified transformer for pedestrian trajectory prediction', in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9675–9684.

[83] K. Li, Y. Chen, M. Shan, J. Li, S. Worrall and E. Nebot, 'Game theory-based simultaneous prediction and planning for autonomous vehicle navigation in crowded environments', in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2023, pp. 2977–2984.

[84] T. Fu, L. Miranda-Moreno and N. Saunier, 'A novel framework to evaluate pedestrian safety at non-signalized locations', *Accident Analysis & Prevention*, vol. 111, pp. 23–33, 2018.

[85] A. Rasouli and I. Kotseruba, 'Intend-wait-cross: Towards modeling realistic pedestrian crossing behavior', in *2022 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2022, pp. 83–90.

[86] K. Tian et al., 'Interacting with yielding vehicles: A perceptually plausible model for pedestrian road crossing decisions', *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[87] J. Pekkanen et al., 'Variable-drift diffusion models of pedestrian road-crossing decisions', *Computational Brain & Behavior*, vol. 5, no. 1, pp. 60–80, 2022.

[88] Q. Liu, X. Li, S. Yuan and Z. Li, 'Decision-making technology for autonomous vehicles: Learning-based methods, applications and future outlook', in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 30–37.

[89] J. Palatti, A. Aksjonov, G. Alcan and V. Kyrki, 'Planning for safe abortable overtaking maneuvers in autonomous driving', in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2021, pp. 508–514.

[90] J. Leonard et al., 'A perception-driven autonomous urban vehicle', *Journal of Field Robotics*, vol. 25, no. 10, pp. 727–774, 2008.

[91] J. Ziegler et al., 'Making bertha drive—an autonomous journey on a historic route', *IEEE Intelligent transportation systems magazine*, vol. 6, no. 2, pp. 8–20, 2014.

[92] T. Alberi et al., 'Odin: Team victortango's entry in the darpa urban challenge',

[93] M. Montemerlo et al., 'Junior: The stanford entry in the urban challenge', *Journal of field Robotics*, vol. 25, no. 9, pp. 569–597, 2008.

[94] E. Onieva, V. Milanés, J. Villagra, J. Pérez and J. Godoy, 'Genetic optimization of a vehicle fuzzy decision system for intersections', *Expert Systems with Applications*, vol. 39, no. 18, pp. 13 148–13 157, 2012.

[95] M. Ammour, R. Orjuela and M. Basset, 'A mpc combined decision making and trajectory planning for autonomous vehicle collision avoidance', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 805–24 817, 2022.

[96] M. A. Abbas, R. Milman and J. M. Eklund, 'Obstacle avoidance in real time with nonlinear model predictive control of autonomous vehicles', *Canadian journal of electrical and computer engineering*, vol. 40, no. 1, pp. 12–22, 2017.

[97] J. Suh, H. Chae and K. Yi, 'Stochastic model-predictive control for lane change decision of automated driving vehicles', *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 4771–4782, 2018.

[98] T. Zhang, J. Zhan, J. Shi, J. Xin and N. Zheng, 'Human-like decision-making of autonomous vehicles in dynamic traffic scenarios', *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 10, pp. 1905–1917, 2023.

[99] L. Le Mero, D. Yi, M. Dianati and A. Mouzakitis, 'A survey on imitation learning techniques for end-to-end autonomous vehicles', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 128–14 147, 2022.

[100] W. Wang, L. Jiang, S. Lin, H. Fang and Q. Meng, 'Imitation learning based decision-making for autonomous vehicle control at traffic roundabouts', *Multimedia Tools and Applications*, vol. 81, no. 28, pp. 39 873–39 889, 2022.

[101] E. Bronstein et al., 'Hierarchical model-based imitation learning for planning in autonomous driving', in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022, pp. 8652–8659.

[102] D. A. Pomerleau, 'Alvinn: An autonomous land vehicle in a neural network', *Advances in neural information processing systems*, vol. 1, 1988.

[103] M. Bojarski et al., 'End to end learning for self-driving cars', *arXiv preprint arXiv:1604.07316*, 2016.

[104] S. Hecker, D. Dai and L. Van Gool, 'End-to-end learning of driving models with surround-view cameras and route planners', in *Proceedings of the european conference on computer vision (eccv)*, 2018, pp. 435–453.

[105] H. Haavaldsen, M. Aasboe and F. Lindseth, 'Autonomous vehicle control: End-to-end learning in simulated urban environments', in *Symposium of the Norwegian AI Society*, Springer, 2019, pp. 40–51.

[106] P. Cai, Y. Sun, Y. Chen and M. Liu, 'Vision-based trajectory planning via imitation learning for autonomous vehicles', in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2019, pp. 2736–2742.

[107] L. Le Mero, D. Yi, M. Dianati and A. Mouzakitis, 'A survey on imitation learning techniques for end-to-end autonomous vehicles', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 128–14 147, 2022.

[108] F. Codevilla, M. Müller, A. López, V. Koltun and A. Dosovitskiy, 'End-to-end driving via conditional imitation learning', in *2018 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2018, pp. 4693–4700.

[109] S. Nozari, A. Krayani, P. Marin-Plaza, L. Marcenaro, D. M. Gomez and C. Regazzoni, 'Active inference integrated with imitation learning for autonomous driving', *IEEE Access*, vol. 10, pp. 49 738–49 756, 2022.

[110] M. Zare, P. M. Kebria, A. Khosravi and S. Nahavandi, 'A survey of imitation learning: Algorithms, recent developments, and challenges', *IEEE Transactions on Cybernetics*, 2024.

[111] P. Wang, D. Liu, J. Chen, H. Li and C.-Y. Chan, 'Decision making for autonomous driving via augmented adversarial inverse reinforcement learning', in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 1036–1042.

[112] A. Y. Ng, S. Russell et al., 'Algorithms for inverse reinforcement learning.', in *Icml*, vol. 1, 2000, p. 2.

[113] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey et al., 'Maximum entropy inverse reinforcement learning.', in *Aaai*, Chicago, IL, USA, vol. 8, 2008, pp. 1433–1438.

[114] M. Wulfmeier, P. Ondruska and I. Posner, 'Maximum entropy deep inverse reinforcement learning', *arXiv preprint arXiv:1507.04888*, 2015.

[115] J. A. Silva, V. Grassi and D. F. Wolf, 'Continuous deep maximum entropy inverse reinforcement learning using online pomdp', in *2019 19th International Conference on Advanced Robotics (ICAR)*, IEEE, 2019, pp. 382–387.

[116] Y. Wang, Y. Niu, M. Xiao, W. Zhu and X. You, 'Modeling framework of human driving behavior based on deep maximum entropy inverse reinforcement learning', *Physica A: Statistical Mechanics and its Applications*, vol. 652, p. 130 052, 2024.

[117] J. Ho and S. Ermon, 'Generative adversarial imitation learning', *Advances in neural information processing systems*, vol. 29, 2016.

[118] M. Ganesan, S. Kandhasamy, B. Chokkalingam and L. Mihet-Popa, 'A comprehensive review on deep learning-based motion planning and end-to-end learning for self-driving vehicle', *IEEE Access*, vol. 12, pp. 66 031–66 067, 2024.

[119] Y. Li, J. Song and S. Ermon, 'Infogail: Interpretable imitation learning from visual demonstrations', *Advances in neural information processing systems*, vol. 30, 2017.

[120] B. R. Kiran et al., 'Deep reinforcement learning for autonomous driving: A survey', *IEEE transactions on intelligent transportation systems*, vol. 23, no. 6, pp. 4909–4926, 2021.

[121] L. García Cuenca, E. Puertas, J. Fernandez Andrés and N. Aliane, 'Autonomous driving in roundabout maneuvers using reinforcement learning with q-learning', *Electronics*, vol. 8, no. 12, p. 1536, 2019.

[122] M. Ahmed, C. P. Lim and S. Nahavandi, 'A deep q-network reinforcement learning-based model for autonomous driving', in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2021, pp. 739–744.

[123] A. Khlifi, M. Othmani and M. Kherallah, 'A novel approach to autonomous driving using double deep q-network-bsed deep reinforcement learning', *World Electric Vehicle Journal*, vol. 16, no. 3, p. 138, 2025.

[124] R. Zhao, Y. Li, Y. Fan, F. Gao, M. Tsukada and Z. Gao, 'A survey on recent advancements in autonomous driving using deep reinforcement learning: Applications, challenges, and solutions', *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[125] R. J. Williams, 'Simple statistical gradient-following algorithms for connectionist reinforcement learning', *Machine learning*, vol. 8, no. 3, pp. 229–256, 1992.

[126] M. Sewak, S. K. Sahay and H. Rathore, 'Policy-approximation based deep reinforcement learning techniques: An overview', *Information and Communication Technology for Competitive Strategies (ICTCS 2020) ICT: Applications and Social Interfaces*, pp. 493–507, 2021.

[127] F. Scharf, F. Helfenstein and J. Jäger, 'Actor vs critic: Learning the policy or learning the value', in *Reinforcement Learning Algorithms: Analysis and Applications*, Springer, 2021, pp. 123–133.

[128] A. Kendall et al., 'Learning to drive in a day', in *2019 international conference on robotics and automation (ICRA)*, IEEE, 2019, pp. 8248–8254.

[129] B. B. Elallid, H. El Alaoui and N. Benamar, 'Deep reinforcement learning for autonomous vehicle intersection navigation', in *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, IEEE, 2023, pp. 308–313.

[130] X. Tang, B. Huang, T. Liu and X. Lin, 'Highway decision-making and motion planning for autonomous driving via soft actor-critic', *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 4706–4717, 2022.

[131] B. Hu, L. Jiang, S. Zhang and Q. Wang, 'An explainable and robust motion planning and control approach for autonomous vehicle on-ramping merging task using deep reinforcement learning', *IEEE Transactions on Transportation Electrification*, vol. 10, no. 3, pp. 6488–6496, 2023.

[132] P. Maramotti, A. P. Capasso, G. Bacchiani and A. Broggi, 'Tackling real-world autonomous driving using deep reinforcement learning', in *2022 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2022, pp. 1274–1281.

[133] F. Yang, X. Li, Q. Liu, Z. Li and X. Gao, 'Generalized single-vehicle-based graph reinforcement learning for decision-making in autonomous driving', *Sensors*, vol. 22, no. 13, p. 4935, 2022.

[134] D. Zhou, P. Hang and J. Sun, 'Reasoning graph-based reinforcement learning to cooperate mixed connected and autonomous traffic at unsignalized intersections', *Transportation Research Part C: Emerging Technologies*, vol. 167, p. 104 807, 2024.

[135] Z. Yang, Z. Wu, Y. Wang and H. Wu, 'Deep reinforcement learning lane-changing decision algorithm for intelligent vehicles combining lstm trajectory prediction', *World Electric Vehicle Journal*, vol. 15, no. 4, p. 173, 2024.

[136] B. Thomas, D. Sanjana, S. Muraleedas, V. V. Nair, N. S. Karat and A. Aprem, 'Autonomous intersection management using gated recurrent units', in *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, IEEE, 2024, pp. 1–6.

[137] H. Liu, Z. Huang, X. Mo and C. Lv, 'Augmenting reinforcement learning with transformer-based scene representation learning for decision-making of autonomous driving', *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 3, pp. 4405–4421, 2024.

[138] G. Chen, Y. Zhang and X. Li, 'Attention-based highway safety planner for autonomous driving via deep reinforcement learning', *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 162–175, 2023.

[139] Y. Gal et al., 'Uncertainty in deep learning', 2016.

[140] K. Wang, C. Shen, X. Li and J. Lu, 'Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications', *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[141] R. Sachdeva, R. Gakhar, S. Awasthi, K. Singh, A. Pandey and A. S. Parihar, 'Uncertainty and noise aware decision making for autonomous vehicles-a bayesian approach', *IEEE Transactions on Vehicular Technology*, 2024.

[142] J. Liu, H. Wang, L. Peng, Z. Cao, D. Yang and J. Li, 'Pnnuad: Perception neural networks uncertainty aware decision-making for autonomous vehicle', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 355–24 368, 2022.

[143] X. Tang et al., 'Prediction-uncertainty-aware decision-making for autonomous vehicles', *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 849–862, 2022.

[144] S. Xu, Q. Liu, Y. Hu, M. Xu and J. Hao, 'Decision-making models on perceptual uncertainty with distributional reinforcement learning', *Green Energy and Intelligent Transportation*, vol. 2, no. 2, p. 100 062, 2023.

[145] J. Dinneweth, A. Boubezoul, R. Mandiau and S. Espié, 'Multi-agent reinforcement learning for autonomous vehicles: A survey', *Autonomous Intelligent Systems*, vol. 2, no. 1, p. 27, 2022.

[146] J. K. Gupta, M. Egorov and M. Kochenderfer, 'Cooperative multi-agent control using deep reinforcement learning', in *International conference on autonomous agents and multiagent systems*, Springer, 2017, pp. 66–83.

[147] D. Chen et al., 'Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic', *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 11 623–11 638, 2023.

[148] Z. Zheng and S. Gu, 'Safe multi-agent reinforcement learning with bilevel optimization in autonomous driving', *IEEE Transactions on Artificial Intelligence*, 2024.

[149] Y. Guo, J. Liu, R. Yu, P. Hang and J. Sun, 'Mappo-pis: A multi-agent proximal policy optimization method with prior intent sharing for cavs' cooperative decision-making', in *European Conference on Computer Vision*, Springer, 2024, pp. 244–263.

[150] Q. Liu, X. Hu and S. Li, 'Research on automatic vehicle lane changing model based on masac-discrete algorithm', in *Sixth International Conference on Advanced Electronic Materials, Computers, and Software Engineering (AEMCSE 2023)*, SPIE, vol. 12787, 2023, pp. 611–620.

[151] L. Canese et al., 'Multi-agent reinforcement learning: A review of challenges and applications', *Applied Sciences*, vol. 11, no. 11, p. 4948, 2021.

[152] A. R. Alozi and M. Hussein, 'How do active road users act around autonomous vehicles? an inverse reinforcement learning approach', *Transportation research part C: emerging technologies*, vol. 161, p. 104 572, 2024.

[153] Z. Zhang et al., 'Decision-making of autonomous vehicles in interactions with jaywalkers: A risk-aware deep reinforcement learning approach', *Accident Analysis & Prevention*, vol. 210, p. 107 843, 2025.

[154] H. Li, J. Huang, Z. Cao, D. Yang and Z. Zhong, 'Stochastic pedestrian avoidance for autonomous vehicles using hybrid reinforcement learning', *Frontiers of Information Technology & Electronic Engineering*, vol. 24, no. 1, pp. 131–140, 2023.

[155] L. Crosato, H. P. Shum, E. S. Ho and C. Wei, 'Interaction-aware decision-making for automated vehicles using social value orientation', *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1339–1349, 2022.

[156] R. Trumpp, H. Bayerlein and D. Gesbert, 'Modeling interactions of autonomous vehicles and pedestrians with deep multi-agent reinforcement learning for collision avoidance', in *2022 IEEE intelligent vehicles symposium (IV)*, IEEE, 2022, pp. 331–336.

[157] W. Hu, H. Mu, Y. Chen, Y. Liu and X. Li, 'Modeling interactions of autonomous/manual vehicles and pedestrians with a multi-agent deep deterministic policy gradient', *Sustainability*, vol. 15, no. 7, p. 6156, 2023.

[158] S. Malik, M. A. Khan, H. El-Sayed, J. Khan and O. Ullah, 'How do autonomous vehicles decide?', *Sensors*, vol. 23, no. 1, p. 317, 2022.

[159] D. Silver, S. Singh, D. Precup and R. S. Sutton, 'Reward is enough', *Artificial intelligence*, vol. 299, p. 103 535, 2021.

[160] H. A. Simon, *Theory of games and economic behavior*. 1945.

[161] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

[162] E. van Damme, 'Game theory: Noncooperative games', in *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier, 2015, pp. 582–591.

[163] D. Novikov, V. Korepanov and A. Chkhartishvili, 'Reflexion in mathematical models of decision-making', *International Journal of Parallel, Emergent and Distributed Systems*, vol. 33, no. 3, pp. 319–335, 2018.

[164] T. Bjørnskau, 'The zebra crossing game–using game theory to explain a discrepancy between road user behaviour and traffic rules', *Safety science*, vol. 92, pp. 298–301, 2017.

[165] C. Fox, F. Camara, G. Markkula, R. Romano, R. Madigan, N. Merat et al., 'When should the chicken cross the road', *Game theory for autonomous vehicle-human interactions*, 2018.

[166] F. Camara, R. Romano, G. Markkula, R. Madigan, N. Merat and C. Fox, 'Empirical game theory of pedestrian interaction for autonomous vehicles', in *Proceedings of measuring behavior 2018*, Manchester Metropolitan University, 2018, pp. 238–244.

[167] X. Chen, Y. Sun, Y. Ou, X. Zheng, Z. Wang and M. Li, 'A conflict decision model based on game theory for intelligent vehicles at urban unsignalized intersections', *IEEE Access*, vol. 8, pp. 189 546–189 555, 2020.

[168] R. E. Amini, A. Dhamaniya and C. Antoniou, 'Towards a game theoretic approach to model pedestrian road crossings', *Transportation research procedia*, vol. 52, pp. 692–699, 2021.

[169] X. Sun, K. Lin, Y. Wang, S. Ma and H. Lu, 'A study on pedestrian–vehicle conflict at unsignalized crosswalks based on game theory', *Sustainability*, vol. 14, no. 13, p. 7652, 2022.

[170] C. M. Harris, 'Autonomous vehicle decision-making: Should we be bio-inspired?', in *Conference Towards Autonomous Robotic Systems*, Springer, 2017, pp. 315–324.

[171] J. K. Goeree and C. A. Holt, 'Ten little treasures of game theory and ten intuitive contradictions', *American Economic Review*, vol. 91, no. 5, pp. 1402–1422, 2001.

[172] P. Chen, C. Wu and S. Zhu, 'Interaction between vehicles and pedestrians at uncontrolled mid-block crosswalks', *Safety science*, vol. 82, pp. 68–76, 2016.

[173] C. Camerer, *Behavioral game theory: Experiments in strategic interaction*. Princeton university press, 2003.

[174] A. H. Kalantari, Y. Yang, Y. M. Lee, N. Merat and G. Markkula, 'Driver-pedestrian interactions at unsignalized crossings are not in line with the nash equilibrium', *IEEE Access*, vol. 11, pp. 110 707–110 723, 2023.

[175] Y. Zhang and J. D. Fricker, 'Incorporating conflict risks in pedestrian-motorist interactions: A game theoretical approach', *Accident Analysis & Prevention*, vol. 159, p. 106 254, 2021.

[176] H. Li, H. Hu, Z. Zhang and Y. Zhang, 'The role of yielding cameras in pedestrian-vehicle interactions at un-signalized crosswalks: An application of game theoretical model', *Transportation research part F: traffic psychology and behaviour*, vol. 92, pp. 27–43, 2023.

[177] D. Levin and L. Zhang, 'Bridging level-k to nash equilibrium', *Review of Economics and Statistics*, vol. 104, no. 6, pp. 1329–1340, 2022.

[178] R. Tian, S. Li, N. Li, I. Kolmanovsky, A. Girard and Y. Yildiz, 'Adaptive game-theoretic decision making for autonomous vehicle control at roundabouts', in *2018 IEEE conference on decision and control (CDC)*, IEEE, 2018, pp. 321–326.

[179] G. S. Sankar and K. Han, 'Adaptive robust game-theoretic decision making strategy for autonomous vehicles in highway', *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 484–14 493, 2020.

[180] R. Tian, N. Li, I. Kolmanovsky, Y. Yildiz and A. R. Girard, 'Game-theoretic modeling of traffic in unsignalized intersection network for autonomous vehicle control verification and validation', *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2211–2226, 2020.

[181] S. Li, N. Li, A. Girard and I. Kolmanovsky, 'Decision making in dynamic and interactive environments based on cognitive hierarchy theory, bayesian inference, and predictive control', in *2019 IEEE 58th Conference on Decision and Control (CDC)*, IEEE, 2019, pp. 2181–2187.

[182] V. V. Unhelkar, C. Pérez-D'Arpino, L. Stirling and J. A. Shah, 'Human-robot co-navigation using anticipatory indicators of human walking motion', in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2015, pp. 6183–6190.

[183] P. Trautman, J. Ma, R. M. Murray and A. Krause, 'Robot navigation in dense human crowds: Statistical models and experimental studies of human–robot cooperation', *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 335–356, 2015.

[184] J. Godoy, I. Karamouzas, S. J. Guy and M. L. Gini, 'Moving in a crowd: Safe and efficient navigation among heterogeneous agents.', in *IJCAI*, 2016, pp. 294–300.

[185] M. Sun, F. Baldini, P. Trautman and T. Murphey, 'Move beyond trajectories: Distribution space coupling for crowd navigation', *arXiv preprint arXiv:2106.13667*, 2021.

[186] H. Fu, Q. Wang and H. He, 'Path-following navigation in crowds with deep reinforcement learning', *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 20 236–20 245, 2024.

[187] T. Ma, Z. Liu, T. Liu, Y. Zhao and Y. Chai, 'A spatiotemporal graphical attention navigation algorithm based on limited state information', *IEEE Transactions on Computational Social Systems*, vol. 11, no. 5, pp. 6407–6421, 2024.

[188] M. Everett, Y. F. Chen and J. P. How, 'Motion planning among dynamic, decision-making agents with deep reinforcement learning', in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 3052–3059.

[189] X. Lu, A. Faragasso, Y. Wang, A. Yamashita and H. Asama, 'Group-aware robot navigation in crowds using spatio-temporal graph attention network with deep reinforcement learning', *IEEE Robotics and Automation Letters*, 2025.

[190] X. Zhou, S. Piao, W. Chi, L. Chen and W. Li, 'Her-drl: Heterogeneous relational deep reinforcement learning for single-robot and multi-robot crowd navigation', *IEEE Robotics and Automation Letters*, 2025.

[191] Y. Cui et al., 'Learning hierarchical graph-based policy for goal-reaching in unknown environments', *IEEE Robotics and Automation Letters*, vol. 9, no. 6, pp. 5655–5662, 2024.

[192] L. Dong, Z. He, C. Song, X. Yuan and H. Zhang, 'Multi-robot social-aware cooperative planning in pedestrian environments using attention-based actor-critic', *Artificial Intelligence Review*, vol. 57, no. 4, p. 108, 2024.

[193] W. Shi, Y. Zhou, X. Zeng, S. Li and M. Bennewitz, 'Enhanced spatial attention graph for motion planning in crowded, partially observable environments', in *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 4750–4756.

[194] Y. Zhou and J. Garcke, 'Learning crowd behaviors in navigation with attention-based spatial-temporal graphs', in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2024, pp. 5485–5491.

[195] H. Zhu, T. Han, W. K. Alhajyaseen, M. Iryo-Asano and H. Nakamura, 'Can automated driving prevent crashes with distracted pedestrians? an exploration of motion planning at unsignalized mid-block crosswalks', *Accident Analysis & Prevention*, vol. 173, p. 106 711, 2022.

[196] K. Tian, G. Markkula, C. Wei and R. Romano, 'Decision model for pedestrian interacting with traffic at uncontrolled intersections', in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2022, pp. 183–188.

[197] I. D. Mienye and Y. Sun, 'A survey of ensemble learning: Concepts, algorithms, applications, and prospects', *Ieee Access*, vol. 10, pp. 99 129–99 149, 2022.

[198] G. Ngo, R. Beard and R. Chandra, 'Evolutionary bagging for ensemble learning', *Neurocomputing*, vol. 510, pp. 1–14, 2022.

[199] Y. Ding, H. Zhu, R. Chen and R. Li, 'An efficient adaboost algorithm with the multiple thresholds classification', *Applied sciences*, vol. 12, no. 12, p. 5872, 2022.

[200] Y. Zhang et al., 'Research and application of adaboost algorithm based on svm', in *2019 IEEE 8th joint international information technology and artificial intelligence conference (ITAIC)*, IEEE, 2019, pp. 662–666.

[201] J. H. Friedman, 'Greedy function approximation: A gradient boosting machine', *Annals of statistics*, pp. 1189–1232, 2001.

[202] M. Angelini, G. Blasilli, S. Lenti and G. Santucci, 'A visual analytics conceptual framework for explorable and steerable partial dependence analysis', *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 8, pp. 4497–4513, 2023.

[203] P. Biecek and T. Burzykowski, *Explanatory model analysis: explore, explain, and examine predictive models*. Chapman and Hall/CRC, 2021.

[204] Q. Zhao and T. Hastie, 'Causal interpretations of black-box models', *Journal of Business & Economic Statistics*, vol. 39, no. 1, pp. 272–281, 2021.

[205] M. Liu, G. Lu, Y. Wang and Z. Zhang, 'Analyzing drivers' crossing decisions at unsignalized intersections in china', *Transportation research part F: traffic psychology and behaviour*, vol. 24, pp. 244–255, 2014.

[206] P. L. Olson, 'Driver perception response time', *SAE transactions*, pp. 851–861, 1989.

[207] X. Zhao, H. Yang, Y. Yao, H. Qi, M. Guo and Y. Su, 'Factors affecting traffic risks on bridge sections of freeways based on partial dependence plots', *Physica A: Statistical Mechanics and its Applications*, vol. 598, p. 127 343, 2022.

[208] Y. Yao, X. Zhao, Y. Wu, Y. Zhang and J. Rong, 'Clustering driver behavior using dynamic time warping and hidden markov model', *Journal of Intelligent Transportation Systems*, vol. 25, no. 3, pp. 249–262, 2021.

[209] Y. Chen, K. Wang and J. J. Lu, 'Feature selection for driving style and skill clustering using naturalistic driving data and driving behavior questionnaire', *Accident Analysis & Prevention*, vol. 185, p. 107 022, 2023.

[210] K. Tian et al., 'Deconstructing pedestrian crossing decisions in interactions with continuous traffic: An anthropomorphic model', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 2466–2478, 2023.

[211] S. Rezwana and N. Lownes, 'Interactions and behaviors of pedestrians with autonomous vehicles: A synthesis', *Future Transportation*, vol. 4, no. 3, pp. 722–745, 2024.

[212] P. Pouya and A. M. Madni, 'Expandable-partially observable markov decision-process framework for modeling and analysis of autonomous vehicle behavior', *IEEE Systems Journal*, vol. 15, no. 3, pp. 3714–3725, 2020.

[213] H. Bai, S. Cai, N. Ye, D. Hsu and W. S. Lee, 'Intention-aware online pomdp planning for autonomous driving in a crowd', in *2015 ieee international conference on robotics and automation (icra)*, IEEE, 2015, pp. 454–460.

[214] Y.-C. Hsu, S. Gopalswamy, S. Saripalli and D. A. Shell, 'A pomdp treatment of vehicle-pedestrian interaction: Implicit coordination via uncertainty-aware planning', in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020, pp. 1984–1991.

[215] D. Zhu, N. Sze, Z. Feng and Z. Yang, 'A two-stage safety evaluation model for the red light running behaviour of pedestrians using the game theory', *Safety science*, vol. 147, p. 105 600, 2022.

[216] L. Peters et al., 'Contingency games for multi-agent interaction', *IEEE Robotics and Automation Letters*, vol. 9, no. 3, pp. 2208–2215, 2024.

[217] J. R. Wright and K. Leyton-Brown, 'Level-0 meta-models for predicting human behavior in games', in *Proceedings of the fifteenth ACM conference on Economics and computation*, 2014, pp. 857–874.

[218] F. M. Shiri, T. Perumal, N. Mustapha and R. Mohamed, 'A comprehensive overview and comparative analysis on deep learning models: Cnn, rnn, lstm, gru', *arXiv preprint arXiv:2305.17473*, 2023.

[219] T. Qie, W. Wang, C. Yang and Y. Li, 'A self-trajectory prediction approach for autonomous vehicles using distributed decouple lstm', *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 6708–6717, 2024.

[220] S. Das, A. Tariq, T. Santos, S. S. Kantareddy and I. Banerjee, 'Recurrent neural networks (rnns): Architectures, training tricks, and introduction to influential research', *Machine learning for Brain disorders*, pp. 117–138, 2023.

[221] W. Li and K. E. Law, 'Deep learning models for time series forecasting: A review', *IEEE Access*, vol. 12, pp. 92 306–92 327, 2024.

[222] M. Dang, Y. Jin, P. Hang, L. Crosato, Y. Sun and C. Wei, 'Coupling intention and actions of vehicle–pedestrian interaction: A virtual reality experiment study', *Accident Analysis & Prevention*, vol. 203, p. 107 639, 2024.

[223] M. A. Costa-Gomes, V. P. Crawford and N. Iriberri, 'Comparing models of strategic thinking in van huyck, battalio, and beil's coordination games', *Journal of the European Economic Association*, vol. 7, no. 2-3, pp. 365–376, 2009.

[224] A. Salmerón, R. Rumí, H. Langseth, T. D. Nielsen and A. L. Madsen, 'A review of inference algorithms for hybrid bayesian networks', *Journal of Artificial Intelligence Research*, vol. 62, pp. 799–828, 2018.

[225] M. H. DeGroot, *Optimal statistical decisions*. John Wiley & Sons, 2005.

[226] Q. Guo and P. J. Gmytrasiewicz, 'Modeling bounded rationality of agents during interactions.', in *Interactive Decision Theory and Game Theory*, 2011.

[227] M. Świechowski, K. Godlewski, B. Sawicki and J. Mańdziuk, 'Monte carlo tree search: A review of recent modifications and applications', *Artificial Intelligence Review*, vol. 56, no. 3, pp. 2497–2562, 2023.

[228] L. Kocsis and C. Szepesvári, 'Bandit based monte-carlo planning', in *European conference on machine learning*, Springer, 2006, pp. 282–293.

[229] E. Ejichukwu, L. Tong, G. Hazime and B. Jia, 'Enhancing autonomous vehicle design and testing: A comprehensive review of ar and vr integration', *arXiv preprint arXiv:2404.19021*, 2024.

[230] D. Yang, K. Redmill and Ü. Özgüner, 'A multi-state social force based framework for vehicle-pedestrian interaction in uncontrolled pedestrian crossing scenarios', in *2020 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2020, pp. 1807–1812.

[231] K. Ryu and N. Mehr, 'Integrating predictive motion uncertainties with distributionally robust risk-aware control for safe robot navigation in crowds', in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2024, pp. 2410–2417.

[232] S. Samavi, J. R. Han, F. Shkurti and A. P. Schoellig, 'Sicnav: Safe and interactive crowd navigation using model predictive control and bilevel optimization', *IEEE Transactions on Robotics*, 2024.

[233] S. Neupane et al., 'Security considerations in ai-robotics: A survey of current methods, challenges, and opportunities', *IEEE Access*, vol. 12, pp. 22 072–22 097, 2024.

[234] M. Kabtoul, A. Spalanzani and P. Martinet, 'Towards proactive navigation: A pedestrian-vehicle cooperation based behavioral model', in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 6958–6964.

[235] J. Y. Angela and P. Dayan, 'Uncertainty, neuromodulation, and attention', *Neuron*, vol. 46, no. 4, pp. 681–692, 2005.

[236] B. Enke and T. Graeber, 'Cognitive uncertainty', *The Quarterly Journal of Economics*, vol. 138, no. 4, pp. 2021–2067, 2023.

[237] A. Pouget, J. Drugowitsch and A. Kepecs, 'Confidence and certainty: Distinct probabilistic quantities for different goals', *Nature neuroscience*, vol. 19, no. 3, pp. 366–374, 2016.

[238] Z. Zhang and F. Xu, 'An overview of the free energy principle and related research', *Neural Computation*, vol. 36, no. 5, pp. 963–1021, 2024.

[239] K. Friston, 'The free-energy principle: A unified brain theory?', *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[240] M. Walters, R. Kaufmann, J. Sefas and T. Kopinski, 'Free energy risk metrics for systemically safe ai: Gatekeeping multi-agent study', *arXiv preprint arXiv:2502.04249*, 2025.

[241] T. Van de Maele, T. Verbelen, O. Çatal, C. De Boom and B. Dhoedt, 'Active vision for robot manipulators using the free energy principle', *Frontiers in neurorobotics*, vol. 15, p. 642 780, 2021.

[242] T. Haarnoja, A. Zhou, P. Abbeel and S. Levine, 'Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor', in *International conference on machine learning*, Pmlr, 2018, pp. 1861–1870.

[243] B. Jiang, Z. Zhang, D. Lin, J. Tang and B. Luo, 'Semi-supervised learning with graph learning-convolutional networks', in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 11 313–11 320.

[244] N. Uhlemann, F. Fent and M. Lienkamp, 'Evaluating pedestrian trajectory prediction methods with respect to autonomous driving', *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 13 937–13 946, 2024.

[245] M. Dang, D. Zhao, Y. Wang and C. Wei, 'Dynamic game-theoretical decision-making framework for vehicle-pedestrian interaction with human bounded rationality', *IEEE Transactions on Intelligent Transportation Systems*, 2025.