# Human Behavior Modeling and Human Behavior-aware Control of Automated Vehicles for Trustworthy Navigation

by

Suresh Kumaar Jayaraman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in The University of Michigan
2021

Doctoral Committee:

Professor Dawn Tilbury, Chair
Professor Ilya Kolmanovsky
Associate Professor Lionel Robert
Associate Professor Ram Vasudevan
Assistant Professor X. Jessie Yang

Suresh Kumaar Jayaraman
jskumaar@umich.edu
ORCID iD: 0000-0002-4874-8582

*Dedicated to my parents, Jayaraman Radhakrishnan and Sujatha Jayaraman, and my wife, Nivedita Saravanan.*

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Dawn Tilbury, for her constant guidance and encouragement. She has been extremely helpful in developing my technical, research, and personal skills and immensely supportive of my ideas and endeavors throughout graduate school. She has been an excellent mentor and has always been there in my time of need, encouraging and boosting my confidence when I needed them the most. I would like to specially thank my committee members and collaborators, Professors Lionel Robert and Jessie Yang, for their support and encouragement, right from the start of my graduate program. The multi-disciplinary nature of the research initiated by these three Professors is what first drew me towards pursuing a Ph.D. I would also like to thank my other committee members Professors Ilya Kolmanovsky and Ram Vasudevan, for providing their support and feedback that improved the dissertation.

I would like to thank the Department of Mechanical Engineering, Rackham Graduate School, and the University of Michigan for giving me the opportunity to pursue the doctoral degree and providing financial support during my time at the university. In addition, I would like to thank the Toyota Research Institute and the Automotive Research Center for providing financial assistance.

I really appreciate the support I received from the MAVRIC lab members. The multi-disciplinary culture and environment that the Professors have fostered in the MAVRIC lab have deeply broadened my perspectives. Specifically, I would like to thank Hebert Azevedo-Sa. He is usually the first person I discuss my ideas with and has been an excellent critique. I would also like to thank Connor Esterwood, Na Du, Qiaoning Zhang, and Huajing Zhao for the numerous discussions and help with my user studies; especially Connor, who took on a variety of roles to help with my user study—from an engineer to a tailor, to even a hidden driver.

Outside of the University of Michigan, I would like to thank my undergraduate advisor, Professor Madhu M., and my internship advisor at the Indian Institute of Technology-Madras, Professor Saravanan Gurunathan. They encouraged me to pursue research and provided me with the necessary opportunities. A special thanks

to SajaySurya Ganesh, a close friend and collaborator in my early research projects, with who I discuss ideas even now.

Last but not least, I would like to thank my family and friends for supporting me during the past several years. My friends at Ann Arbor made life away from home much easier; they are like my second family. A long list of people from my Master's and Ph.D. programs at the University of Michigan has played an essential role in my graduate experience. Still, I would like to especially thank Sandipp Krishnan Ravi, Subramaniam Balakrishna, Rahasudha Kannan, and Paavai Pari for all their love and support. I will fondly remember my time at the University of Michigan and in Ann Arbor because of all of the people I encountered, the friends I made, and the experiences I had. My parents, wife, and extended family have all been incredibly supportive of the pursuit of my degree, and I am eternally grateful for their love and guidance.

# TABLE OF CONTENTS

# LIST OF FIGURES

**Figure**

# LIST OF TABLES

**Table**

# LIST OF ABBREVIATIONS

**ADE** average displacement error . . . . . 57
**AOI** areas of interest . . . . . 31
**AV** automated vehicle . . . . . 1
**AVIVR** automated vehicle interaction in virtual reality . . . . . 57
**B-MPC** behavior-aware model predictive controller . . . . . 107
**EGT** expectation of ground truth . . . . . 96
**FDE** final displacement error . . . . . 57
**FRS** forward reachable set . . . . . 96
**HDV** human-driven vehicle . . . . . 4
**IHP** interacting hybrid pedestrian . . . . . 83
**IVE** immersive virtual environment . . . . . 25
**MLRM** mixed linear repeated modeling . . . . . 31
**POI** pedestrian of interest . . . . . 9
**SAE** Society of Automotive Engineering . . . . . 2
**WOZ** Wizard-of-Oz . . . . . 13

# ABSTRACT

Robots are increasingly being developed for and used in human-centered environments, and there is a growing need to develop methods for safe human-robot interaction. Automated vehicles (AVs), a type of robot that has garnered much attention recently, can increase transport safety, efficiency, and accessibility. However, to realize these benefits, they need to be trusted and accepted by the general public.

This dissertation focuses on the interaction between AVs and pedestrians, one of the most vulnerable types of road users. AVs are a novel technology, and thus the interaction dynamics between AVs and pedestrians are not clear. A major operational challenge for AVs is safe navigation in urban environments around pedestrians. To improve pedestrian trust in AVs, researchers typically develop motion planning methods that can guarantee safe operation. However, in addition to safety, other factors (environmental and behavioral) can influence trust in the AVs and, in turn, their acceptance.

Typically, AVs employ a receding-horizon planning methodology, where they plan for a short horizon ($1 - 2\,s$) while incorporating predictions of pedestrian trajectories to avoid potential collisions. The urban traffic environment is dynamic in nature and constantly changes with the location and the behaviors of the surrounding road agents. This requires the AV to plan in real time. The prediction and planning models, therefore, should be computationally efficient to enable real-time planning. A key challenge in AV planning is balancing safety and performance. Focusing only on safety can lead to highly conservative AV behaviors that are undesirable. Further, to gain public trust and acceptance, the AVs should demonstrate navigation capabilities that are both safe and trustworthy. Extending the prediction and planning horizons to a longer term ($> 5\,s$) could aid the AV in developing such safe and trustworthy trajectories.

This dissertation addresses two high-level research problems in the context of pedestrian-AV interaction—(i) how to predict long-term pedestrian behaviors efficiently and (ii) how to use the pedestrian behavior predictions to plan safe and trustworthy trajectories in real time. This dissertation has four primary contributions.

First, this dissertation characterizes the effects of AV driving behavior and environmental factors on pedestrian's trust in the AVs and pedestrian behavior, based on user studies developed in virtual and controlled real-world environments. Second, a new modeling framework for urban pedestrian behavior based on hybrid systems theory is presented. The framework models the high-level intent and decision-making process of pedestrians and uses a simple continuous motion model. Third, the framework is extended to include interaction between other pedestrians and predict multimodal pedestrian behaviors. The proposed framework is tested on publicly available real-world datasets and a virtual reality dataset collected from a user study. The results show the model's ability to predict long-term multimodal pedestrian behaviors that are intuitive and explainable. Finally, a receding-horizon planner that incorporates the pedestrian predictions is presented. The planner was tested in a simulated traffic environment. Results indicate the potential of the approach to developing safe AV behaviors that are understandable and trustworthy. The models and methods discussed in this dissertation enable a better understanding of human and robot behaviors, thereby aiding in realizing safe and trustworthy human-robot interactions.

# CHAPTER I

# Introduction

This dissertation addresses safe and trustworthy human-robot interaction, where the robots are automated vehicle (AV)s, and the humans are pedestrians. Trustworthy navigation of the AVs results in improved coordination among road agents and improved traffic safety. This dissertation characterizes the relationships between pedestrian behavior, AV behavior, and pedestrian trust in the AVs and develops models of pedestrian and AV behaviors for trustworthy navigation. Current pedestrian behavior models are either limited to the short term, as in the case of dynamics-based models, or are computationally expensive, require large datasets, and lack intuition, as in deep learning models. This research aims to develop computationally efficient long-term pedestrian behavior models and behavior-aware AV controllers that improve trust in AVs and traffic safety.

## 1.1 Motivation

Robotic technology is progressing at a rapid rate, both in terms of reach and intelligence. Robots have started to permeate several areas of humans' lives, such as workplaces, streets, and even our homes, effectively interweaving into our social fabric. This rapid progress of robotic technology widens the gap between the complexity of the technology and human understanding of such technology. Thus robots have to be designed considering their potential interactions with the humans in their interacting environments.

One of the most impactful applications of robotics is automated vehicles (AVs). While AVs have been discussed since the 1950s (refer Figure 1.1) and researched since the 1980s [2], advances in sensing, control, and artificial intelligence have now brought them closer to becoming a reality. AVs potentially have several benefits, such as reduced fossil fuel consumption, increased comfort, and increased connectivity and

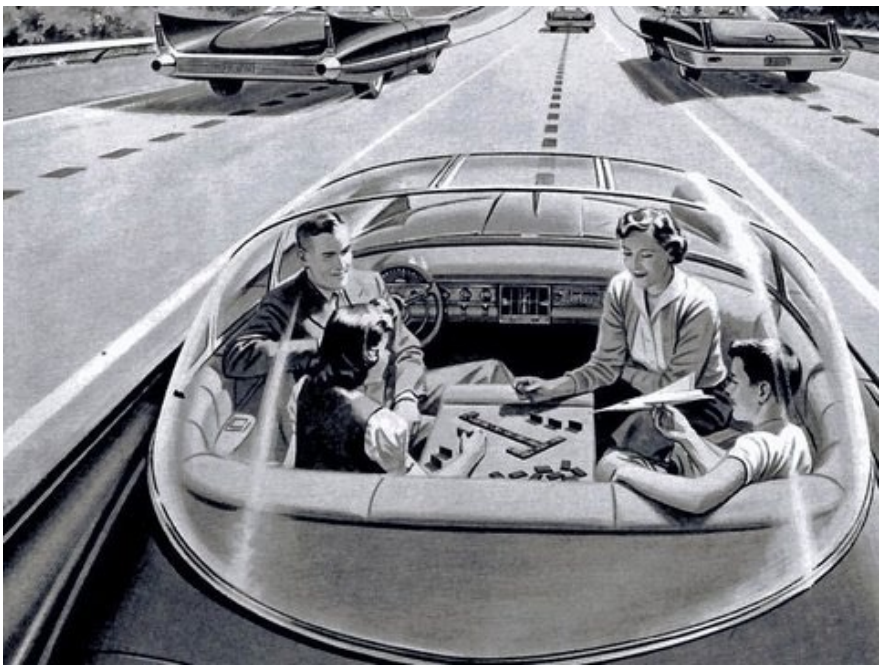


Figure 1.1: A 1950s advertisement featuring automated vehicles [1].

access to transportation. However, the most important benefit is their potential to improve road safety [3]. Existing research estimates that over 90 % of severe motor crashes are caused due to human errors [4]. By effectively augmenting human capabilities (Society of Automotive Engineering (SAE) Level 3 AV, refer Figure 1.2) or eliminating a human driver's necessity (Level 4 and 5 AVs), AVs can potentially reduce crashes and improve traffic safety.

Fagnant and Kockelman [5] predict that a 90% market penetration of AVs will lead to 21,700 lives saved per year and 4.2 million fewer crashes. Therefore, widespread acceptance of AV technology is required to realize these benefits.

Fully self-driving AV technology is still years away from being deployed on public roads. This is partly due to technological limitations, especially in terms of understanding the social behaviors of other road users, and partly due to lack of public acceptance [3]. A major barrier to widespread AV acceptance is public skepticism over the safety of AVs [6, 7]. Incidents such as the Tesla car crashes [8] and the Uber crash [9] involving a pedestrian further impact public perception of AVs. Researchers have identified *social trust* as an important factor to improve the acceptance of AVs by the general public [6]. AVs must therefore demonstrate capabilities for safe and trustworthy navigation to gain public trust and acceptance. Recent research efforts

have focused on developing provably safe navigation algorithms [10, 11] to improve trust and acceptance.

The safety of humans interacting with AVs is critical due to the high risk of collisions. Human interaction with AVs is complicated as various kinds of humans potentially interact with AVs. These humans can be inside the vehicle, such as drivers or passengers, or outside, such as other human drivers, bicyclists, or pedestrians. In this dissertation, we focus on the interaction between pedestrians and AVs for two reasons. First, pedestrians are one of the most vulnerable road users as they are more prone to injuries than other road users in case of a crash. Second, AV interaction with pedestrians is very different from its interaction with drivers/passengers inside the vehicle. Pedestrians are more likely to be less familiar with the novel AV technology as they are not using it or traveling in it. Thus, it is not surprising that most pedestrians are not comfortable sharing the road with AVs [12].

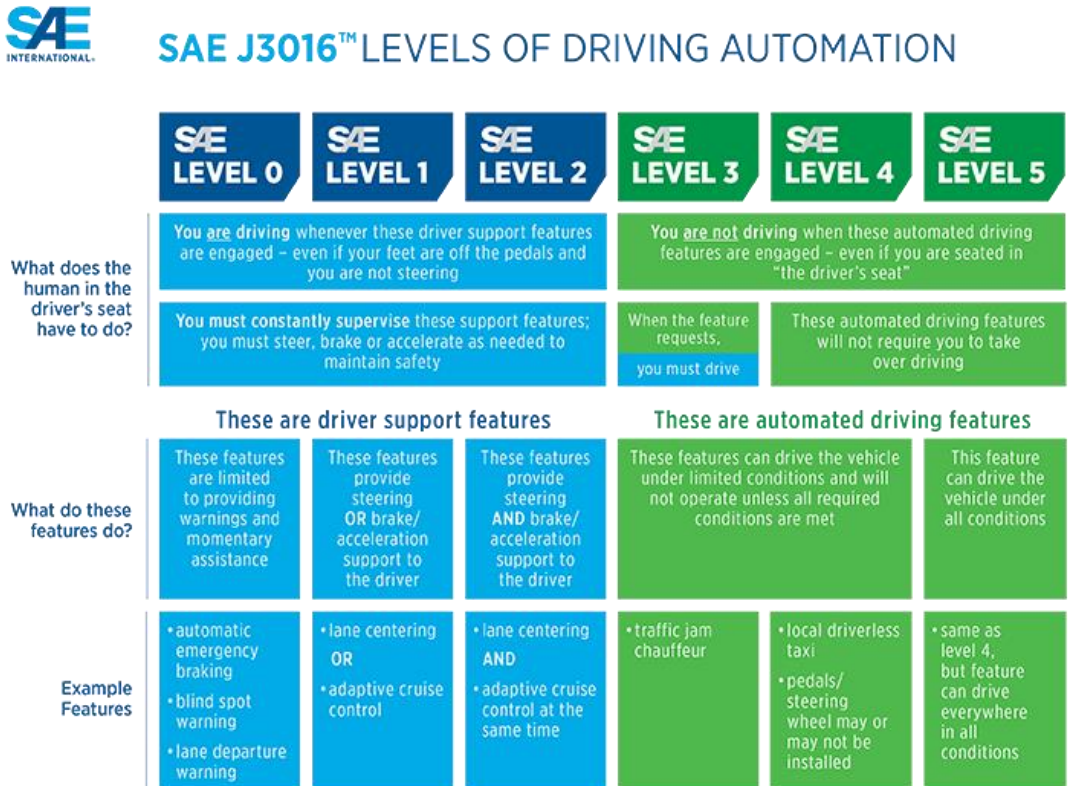



Figure 1.2: SAE levels of driving automation [13].

In the context of this dissertation, *Pedestrian behavior* refers to the actions and, in turn, the paths (or trajectories) taken by pedestrians. Our current understanding of pedestrian behavior may not be directly applicable to their interaction with AVs. The novelty of AV technology warrants a detailed investigation of pedestrian behavior

around AVs.

For safe and trustworthy navigation, AVs must predict the future behavior of pedestrians accurately and react accordingly. Pedestrian behavior depends on several factors, primarily the behavior of the interacting vehicle and the behaviors of nearby pedestrians. Understanding the effects of these different factors enables the accurate prediction of pedestrian behaviors.

The ability to predict pedestrian behavior gives the AVs sufficient time to plan trajectories that are safe and perceived as trustworthy by the pedestrians. The longer the AVs can accurately predict pedestrian behavior, the easier it is to plan safe trajectories. Currently, purely data-driven methods are employed to predict pedestrian behaviors. While the accuracy of such methods is impressive, the lack of intuition makes it difficult for the AVs to interpret the predictions [14]. Currently, AVs in the real world maximize safety and drive conservatively. Such conservative behavior could result in slow traffic flow in addition to garnering unexpected reactions from the public ranging from curiosity to verbal abuse, and vandalism [15]. One of the reasons for such public reactions could be low trust in the AVs.

This dissertation aims to characterize pedestrian-AV interactions and develop methods for safe and trustworthy AV navigation. Particularly, this dissertation revolves around answering four high-level research questions: (i) how does AV driving behavior affect pedestrian trust in the AVs and pedestrian behaviors? (ii) how can we efficiently predict long-term behaviors of pedestrians that are intuitive and explainable? (iii) how to effectively incorporate the interaction between pedestrians to predict their long-term behaviors? (iv) how to use the pedestrian behavior predictions to design AV controllers that improve safety and pedestrian trust?

### 1.1.1 Pedestrians' trust in the AVs

Pedestrian behavior depends on their interaction with oncoming vehicles as well as other pedestrians [14]. Pedestrian interaction with human-driven vehicles (HDVs) is well studied in the literature [16, 17]. However, pedestrian interaction with AVs is relatively less explored.

Pedestrians typically engage in non-verbal communication with the human driver to understand the driver's intent and communicate their intent. For example, pedestrians may use the HDV's motion cues or a head nod from the driver to understand the driver's intent and use non-verbal cues such as gaze or hand gestures to communicate their intent. However, in the case of pedestrian interaction with fully automated vehicles, such modes of communication are different at best and non-existent at worst,

which could cause the pedestrians to rely more on the AV's behavior (motion cues) to gather information about the AV's intent [18, 19].

Pedestrians have to deduce AV intent based on the limited information available to them from AV behavior. Thus there could be uncertainty surrounding the deduced AV intent. Information uncertainty is known to reduce trust in human-human relationships [20]. However, the effects of such information uncertainty on the trust relationships in pedestrian-AV interactions are unknown and are a significant research gap. Particularly, it is unclear how pedestrians would trust the AVs and how their behavior would change with their trust under different situations that vary in the uncertainty of the AV intent. Furthermore, compared to pedestrian interaction with HDVs, the differences in communication may change the way pedestrians interact with AVs. However, it is still unclear how pedestrians' trust changes around AVs compared to their traditional behavior around HDVs, which is another research gap.

The first part of this dissertation addresses these two research gaps by examining the effects of driving behavior and vehicle type on pedestrians' trust and behaviors. Chapter III discusses two user studies that evaluate pedestrian-AV interactions in virtual and real-world environments to examine the relationship between vehicle behavior, pedestrian trust, and pedestrian behavior.

### 1.1.2 Long-term pedestrian behavior prediction

Most prior research has focused on developing prediction algorithms for short-term pedestrian behavior, typically 1-2 $s$ [21–23]. Such short-term predictions may be sufficient to avoid hard collisions at low vehicle speeds ($< 20\,mph$), such as in shared spaces [24]. However, at higher speeds, as in structured urban environments, AVs must predict pedestrian trajectories over a longer term ($> 5\,s$) for safe and comfortable navigation [14, 25]. Existing state-of-the-art methods are data-intensive and computationally expensive. This limits their applicability to complex scenarios where the number of road agents is large, increasing the computational complexity. A lack of computationally efficient methods to predict long-term pedestrian behavior continues to be a research gap.

Trajectory prediction is challenging because of the inherent uncertainty in pedestrian behavior [26, 27]. Their intentions and decisions drive pedestrian behaviors and, in turn, the trajectories traversed by pedestrians. Pedestrian behavior is multimodal in that pedestrians can take multiple possible trajectories in any given situation. An AV can never be entirely sure what a pedestrian will do next; therefore, trajectory prediction is inherently stochastic. On the one hand, it is essential to consider mul-

tiple possible pedestrian behaviors. On the other hand, not all possible pedestrian behaviors are relevant to AVs. For example, AVs should be more concerned about pedestrians who intend to cross and are about to cross than those just walking on the sidewalk. However, most current pedestrian models either produce only a single prediction output [21] which does not capture the uncertainty in pedestrian behaviors, or produce a distribution of behaviors that are arbitrarily sampled [28] lacking any underlying intuition for the sampled behaviors. This lack of intuition of the predictions limits their application to new unseen environments making the models less generalizable. In this dissertation, we develop intuitive methods of pedestrian behavior prediction. Chapter IV focuses on the development of an intuitive and computationally efficient pedestrian behavior model that is suitable for long-term multimodal predictions.

### 1.1.3 Pedestrian-pedestrian interactions in behavior prediction

Another critical facet of pedestrian behavior is their interaction with nearby pedestrians. Pedestrian behavior can be influenced by the actions of nearby pedestrians [16]. However, not all nearby pedestrians influence the behavior of a pedestrian. For example, a pedestrian might be more comfortable in crossing a street after seeing another pedestrian crossing from the opposite side, while the actions of another more closer pedestrian simply walking away from the crosswalk may not be very relevant. Mainly, pedestrians who walk in groups are known to have different behavior than those crossing individually. Generally, a group of pedestrians is found to have more risk-taking behaviors than an individual pedestrian [16]. In Chapter V, we examine the interaction effect with neighboring pedestrians in large-scale datasets and integrate the interaction effect into our pedestrian prediction framework.

### 1.1.4 AV planning for safe and trustworthy navigation

AVs use the predicted pedestrian trajectories for motion planning in real-time. Safety is a priority while planning AV motion. However, the AVs also have other goals such as ensuring they reach the destination in a reasonable amount of time, maintaining ride comfort, etc., while adhering to the traffic laws. AVs develop safe paths that satisfy their objectives—reach the target destination, maintain a comfortable ride, maintain efficient fuel consumption, etc.,—while avoiding collision with the possible future trajectories of the pedestrians [29, 30]. However, the interactions between AVs and pedestrians in urban scenarios are highly dynamic. New pedestrians

can suddenly enter the sidewalks from buildings, corners, or other blind spots. Thus there is no guarantee that a previous path planned by the AV continues to be safe. AVs frequently replan their paths to avoid any potential collisions in such dynamic situations, and the rate of replanning must be high enough to compensate for the dynamically varying environment [31]. Thus the pedestrian prediction algorithms should be computationally efficient to support fast replanning and real-time AV operation. Chapter VI discusses the development of a behavior-aware control algorithm for AV motion planning and methods to design AV behavior for safe and trustworthy navigation around pedestrians.

## 1.2 Scope of the dissertation

A major operational challenge for AVs is safe navigation in urban environments around pedestrians who can change their actions instantaneously. In this dissertation, we focus on the interactions between pedestrians and AVs in urban environments.

We mainly focus on urban pedestrian behavior around unsignalized crosswalks, which are more complex than signalized crosswalks as the right-of-way is unclear [32]. In addition, this dissertation focuses on pedestrian interactions with fully automated vehicles (Levels 4 and 5, refer Figure 1.2) that do not require a backup driver to take over control of the vehicle.

Broadly, there are two categories of safety — provable safety and perceived safety. Provable safety is the actual safety that the AV can achieve, and perceived safety is how safe the humans interacting with the AVs perceive them to be. Though these two are similar, they are not the same. For example, a provably safe navigation plan is to go as close to the pedestrian as safely possible and hard brake every time there is a pedestrian nearby. However, such a hard braking action may not be perceived safe by the pedestrian or the passengers of the AV. Ideally, the AVs need to be both provably safe and perceived to be safe.

Perceived safety can be considered to be more relevant for HRI applications as the behavior of the humans depend on how safe they perceive the robots (in this case, the AVs) to be. This dissertation primarily focuses on perceived safety by developing controllers that aim to (i) avoid collisions and (ii) increase pedestrian-vehicle distance.

## 1.3 Contributions

This section describes the four primary contributions of this dissertation. Each primary contribution corresponds to a chapter in the dissertation. The contributions are named *C1-C4* and referenced accordingly.

1. *C1: Characterization of how driving behavior and vehicle type affect pedestrian's trust.*

   As discussed in Section 1.1.1, the differences in communication and uncertainty of information from AV driving behavior during pedestrian interactions with AVs can result in differences in pedestrian behavior, moderated by their trust. **The first main contribution** of this dissertation is the characterization of the role of driving behavior and vehicle type (HDV or AV) on pedestrian trust and pedestrian behaviors. The contribution is represented by the conclusions of two user studies which are described in Chapter III. The findings of the first user study are published in [33, 34] and can be summarized as follows. In addition to confirming existing research that AV intent can be deduced from their behaviors [18], we found that pedestrian trust and behavior are influenced by AV driving behaviors. We found that pedestrians engaged in observable trusting behaviors such as reduced gaze at AVs, increased walking speed, and increased waiting time before crossing when they reported trusting the AVs. However, the influence of AV driving behavior on pedestrian trust and behavior is valid only at unsignalized crosswalks. From the second user study, we found that pedestrians have different expectations towards AVs than HDVs, resulting in different pedestrian trust for AVs compared to HDVs. We also found that pedestrians' trust in the AVs increases with each subsequent interaction. These findings from the user studies enable the development of pedestrian behavior models and AV control algorithms which are the subsequent contributions in this dissertation.

2. *C2: Development of an explainable and computationally efficient long-term model for multimodal pedestrian behavior.*

   Existing approaches for long-term pedestrian behavior prediction are computationally expensive and require large amounts of labeled data. Further, the multimodal predictions in such approaches lack intuition and explainability. **The second contribution** of this dissertation is the development of an explainable and computationally efficient pedestrian behavior model that is suit-

able for long-term multimodal predictions. The development and validation of a pedestrian behavior model based on hybrid automata theory incorporating pedestrian interaction with an approaching vehicle are explained in Chapter IV and in [35, 36]. We demonstrate the validity of the model on two datasets — one collected from a VR study discussed in Section 3.2.3.4 and the other is an existing dataset that captured pedestrian-HDV interactions in the real-world [37].

3. *C3: Extension of pedestrian behavior model incorporating interactions among multiple pedestrians.*

   Current approaches to modeling pedestrian behavior either isolate the interaction between a pedestrian and the vehicle [38] or incorporate interactions with other pedestrians primarily as a means to avoid collisions and neglect the effect of nearby pedestrians on the decision-making process of the pedestrian of interest (POI). Further, these models incorporate all possible interactions between multiple pedestrians and multiple vehicles in a data-driven framework that lack explainability and increase problem complexity [28, 39]. **The third contribution** of this dissertation is the extension of the pedestrian model (*C2*) to incorporate the effects of pedestrian interaction with neighboring pedestrians efficiently. A limiting interaction region is identified for pedestrians beyond which pedestrian behavior is not affected by other pedestrians. Chapter V discusses the extension of the pedestrian prediction framework from Chapter IV to incorporate their interaction with neighboring pedestrians and an approaching vehicle. The extended model was validated on a pre-existing dataset containing real-world interactions between pedestrians and vehicles [40].

4. *C4: Development of a behavior-aware controller*

   Controllers developed for AV planning are mostly conservative or imitate human driving behaviors. However, *C1* states that pedestrian expectations towards AVs are different from their expectations towards HDVs, and thereby, imitating human driving behaviors may not be ideal. Further, conservative driving behaviors, though safe, may result in traffic congestion. **The fourth contribution** of this dissertation is the development of an AV controller for safe and trustworthy motion planning that incorporates the predictions from the pedestrian behavior models. The controller demonstrates the ability to design driving behaviors that the pedestrians can potentially understand, which could facilitate safe and trustworthy navigation. The development of a behavior-aware

controller based on model predictive control (MPC) framework is discussed in Chapter VI and published in [30, 41].

## 1.4 Dissertation overview

The remainder of the dissertation is structured as follows.

*Chapter 2* discusses the necessary background and state-of-the-art research in pedestrian-AV interaction, pedestrian behavior modeling, and AV planning. We identify and summarize the gaps in existing literature in these areas.

*Chapters 3, 4, 5 and 6* discuss the contributions *C1*, *C2*, *C3*, and *C4* respectively as mentioned in Section 1.3.

*Chapter 7* summarizes the dissertation's contributions and discusses the limitations of the dissertation and directions for future work.

# CHAPTER II

# Background

Researchers have been looking into pedestrian modeling and AV planning for several years, and some excellent methods have been developed for both pedestrian modeling and AV planning. Pedestrian interaction with AVs, on the other hand, is a relatively less explored area. We first discuss the background and current research in pedestrian gap acceptance behavior and pedestrian-AV interaction. We then discuss current methods developed for pedestrian modeling, methods for modeling the interaction between pedestrians, and finally, describe the methods developed for interactive and safe AV planning. We summarize the gaps in the existing literature at the end of each section.

## 2.1 Pedestrian gap acceptance

Pedestrian crossing decision in the presence of an oncoming vehicle can be expressed through the theory of gap acceptance [42]. A pedestrian's decision to cross is normally modeled as a traffic gap acceptance problem that identifies the gaps where a pedestrian feels safe to cross [43, 44]. A *traffic gap* is defined as the time taken by the closest vehicle, traveling at its current velocity, to reach the pedestrian's longitudinal position along the road. Pedestrians, when intending to cross the street, typically decide to cross the street by evaluating the safety of the available traffic gap from the closest interacting vehicle [43, 45]. A gap is considered to be accepted when the pedestrian starts crossing during that gap.

Yannis et al. [43] developed a logistic regression model to predict gap acceptance based on pedestrian waiting time, vehicle distance, and age and gender of pedestrians. More recent approaches have used pedestrian's pose, motion, and vehicle behavior to develop Markovian [22] or neural network models [46, 47] that predict the crossing decisions. These models identify whether pedestrians are standing or walking from

their pose (represented by the joint positions of their skeleton model) and predict their future positions based on the continuous motion associated with the identified action. However, these models cannot accurately predict changes in the actions, especially from standing to walking. Thus, these models have limited applicability to crosswalk scenarios as they do not predict crossing decisions of pedestrians waiting (and standing) at the crosswalk. Furthermore, the underlying assumption in these models is that all pedestrians intend to cross, which is not necessarily true. When they do not intend to cross, pedestrians' behavior is different from when they intend to cross.

## 2.2 Pedestrian-AV interaction

Prior research on pedestrians' interaction with human-driven vehicles (HDVs) has highlighted the importance of non-verbal communication during negotiation between drivers and pedestrians [16, 48]. Human drivers engage in non-verbal communication via eye contact, facial expressions, and hand gestures [49, 50], as can be seen in Fig. 2.1. This is often done to communicate the drivers' intent when negotiating the right-of-way with pedestrians [48]. In the absence of a human driver, it is not surprising that pedestrians have expressed concerns over not knowing or understanding the AV's intention [51, 52]. This uncertainty in AV intent can, in turn, lead to pedestrians misinterpreting the AV's behavior and potentially result in dangerous situations.

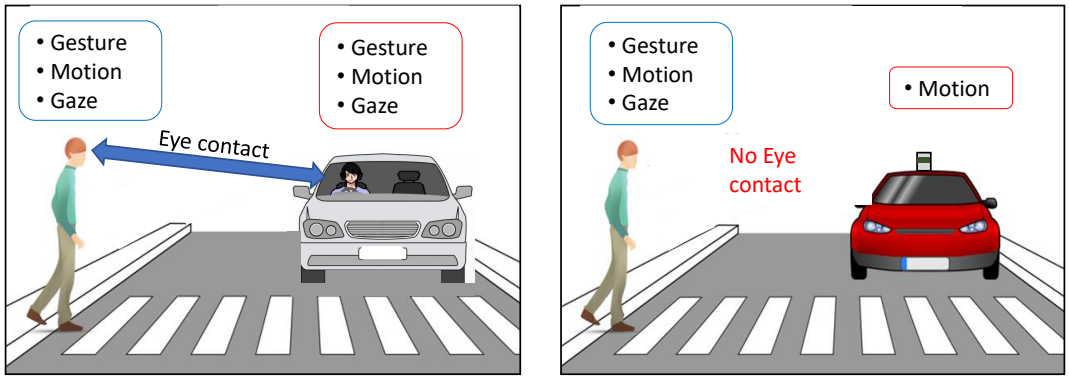



Figure 2.1: A typical interaction of a pedestrian intending to cross a street in the presence of (i) a human-driven vehicle, and (ii) an automated vehicle. The common modes of non-verbal communication between a pedestrian and a human driver are non-existent in the case of an AV.

### 2.2.1 Communication of AV intent

Researchers are thus exploring ways to communicate AV intent to pedestrians. AVs can communicate their intent through explicit or implicit means. Traditional methods of explicit communication in HDVs include indicator lamps, brake lamps, and horns. Current research on AV explicit communication primarily explores the efficacy of additional specialized interfaces such as message boards, LED lights, interactive headlamps, etc., in conveying vehicle intent in the absence of human drivers [53–56]. Although these approaches are valuable and insightful, there is currently no standard communication interface. Moreover, when the number of AVs on the street increases in the future, explicit communication may pose problems to pedestrians, such as information overload.

Implicit rather than explicit communication is a less explored approach to tackling the communication challenge between pedestrians and AVs [57]. Implicit vehicle communication refers to the behavior cues derived from the vehicle's driving [18, 19]. Pedestrians can get information about the AV's intent through its driving behavior, specifically through its motion and kinematics [18, 58]. For example, an AV intending to yield the right-of-way to pedestrians at the crosswalk will do so by starting to slow down, whereas an AV that does not intend to yield will not slow down or may even accelerate.

Research on implicit communications between pedestrians and AVs has focused on the problems with the absence of the human driver. The AV's driving behavior has been used as a form of implicit communication. Typically researchers have varied AV driving behavior from more to less aggressive by varying the vehicle's velocity profile and measured participants' responses to the driving behavior [58, 59]. Studies have shown that AVs can implicitly communicate their intent to pedestrians through their driving behavior [19, 59]. For example, Fuest et al. [19] examined AV intent recognition by pedestrians. They used a Wizard-of-Oz (WOZ) approach where the driver wore a car seat costume and hid in plain sight from the pedestrians. Results indicate that pedestrians, in general, were able to identify the AV's intent of stopping or not stopping from its driving behavior.

The intentions of AVs can also be understood from other contextual elements such as traffic signals. AVs are expected to be much more law-abiding than human drivers [60, 61]. Thus, in situations where the right-of-way is clear, such as at signalized crosswalks, AVs are always expected to follow the traffic rules and stop at the red light. Conversely, in situations where the right-of-way is unclear, pedestrians would be skeptical of AVs. One such situation is an unsignalized crosswalk, where the right-

of-way varies from state to state in the US [32]. In the case of signalized crosswalks, the traffic signal clarifies the right-of-way to all traffic participants. As the AVs are always expected to follow traffic rules, the traffic signal indirectly informs the AVs' intent to the pedestrians. Traffic signals are a part of the infrastructure and dictate the right-of-way. Thus, they can be considered a higher authority, and AVs could be expected to follow the signal irrespective of their driving behavior. Thus the presence of a traffic signal might moderate the effects of a vehicle's driving behavior on pedestrian behavior. This interaction between AV driving behavior and the traffic signal is relatively less explored.

### 2.2.2 Pedestrian trust in AVs

Scholars have also begun examining the impact of AV driving behavior on pedestrian's trust. Pedestrian trust in AVs is highly relevant because pedestrians are warier of crossing in front of an AV without a driver than crossing in front of an HDV [62], indicating less trust in AVs. Thus pedestrians can indirectly express their trust in AVs through their behaviors.

Existing studies varied the AV driving behaviors and explored pedestrian trust through behaviors such as willingness to cross, crossing paths, and comfort of crossing [58, 63, 64]. For example, Rothenbucher et al. [63] explored the reactions of pedestrians upon encountering an AV. They found that people generally crossed the street normally and were tolerant of aggressive driving by the AV. Pedestrians' willingness to cross seemed to be unaffected by the AV's different driving behaviors. However, both Pillai [58] and Zimmermann and Wettach [64] found that when the AV engaged in what would be considered more defensive driving behavior (decelerating early) versus more aggressive driving behavior (decelerating later), they perceived the defensive AV to be more controlled and reliable than the aggressive AV. Overall, there is more evidence that different AV driving behaviors can affect pedestrian trust differently.

Trust in automation studies [65–67] have identified various factors affecting trusting behavior. Generally, trust is only one of the factors that influence behaviors, in addition to workload, situational awareness, system capability, and other contextual and environmental factors [66]. Though trust is related to trusting behaviors, the relationship between trust and trusting behaviors might not be straightforward.

### 2.2.3 Pedestrian behavior around AVs

Current research has compared pedestrian behavioral differences with AVs and HDVs. For example, Habibovic et al. [55] used a WOZ AV and found that pedestrians were less comfortable crossing in front of an AV than an HDV. On the contrary, Rothenbucher et al. [63] and Rodríguez Palmeiro et al. [68] used a WOZ AV and found that people crossed the street similarly in front of AVs and HDVs. Currently, there is no consensus on differences/similarities in pedestrian crossing behavior between AVs and HDVs, which warrants further investigation.

To summarize, there is a lack of understanding of pedestrian behavior during their interaction with AVs, and there is no consensus on differences/similarities in pedestrian crossing behavior between AVs and HDVs. We address these gaps in Chapter III through user studies.

## 2.3 Pedestrian behavior modeling

Understanding pedestrian motion and predicting such motion is crucial for AVs to coexist and interact with pedestrians. As such, pedestrian motion prediction has received significant attention in recent years. The challenge of making accurate predictions of pedestrian motion arises from the complexity of their behavior. A recent survey paper by Rudenko et al. [14] compiled the plethora of research generally in the field of human road agent motion prediction, with the main focus on pedestrians.

Rudenko et al. [14] broadly classified existing methods for predicting pedestrian trajectories into (i) physics-based, (ii) pattern-based, and (iii) planning-based methods. *Physics-based methods* [21, 23, 35, 69, 70] predict by forward simulating a set of explicitly defined dynamics equations that follow a physics-inspired model. *Pattern-based methods* [71–76] approximate an arbitrary dynamics function from the training data. These methods learn pedestrian behaviors by fitting different function approximators (i.e., neural networks, hidden Markov models, Gaussian processes) to the training data. *Planning-based methods* [77–80] explicitly reason about the pedestrian's motion goals and compute policies or path hypotheses that enable an agent to reach those goals. The planning-based methods have been able to predict better the pedestrians' long-term behavior than the other two.

The planning-based methods have demonstrated that using goal locations can improve the long-term prediction capabilities of the models. Goal locations are generally assumed to be known [77]. Alternatively, some works estimate the goal locations by clustering the pedestrian trajectories and identifying the endpoints of each trajectories

cluster [81].

Current state-of-the-art methods model complex pedestrian behaviors through deep-learning approaches [14, 82–84]. Such approaches utilize datasets to learn features that generalize across a variety of scenes. Recurrent neural network (RNN) models, in particular, effectively capture complex agent behaviors and long-term dependencies in road agent behaviors [28, 76, 82, 84–86].

Another approach to model complex pedestrian behaviors is by combining several simple physics-based models into a hybrid system [21, 87–89]. Common approaches include interacting multiple models [70, 90, 91], and dynamic Bayesian networks [21, 92, 93]. Although the physics-based hybrid models require less data than deep learning methods and are more generalizable to new environments, their predictions are limited to the short-term $(1 - 2\,s)$. The discrete states of the pedestrians are usually maneuvers such as walking, running, and stopping [21, 94]. However, it is more relevant for the AVs to predict the decision choices of the pedestrians than their actual maneuvers, which these models do not completely capture. For example, an AV might be more interested in knowing whether a pedestrian walking on the sidewalk will cross or not rather than whether they are walking or running on the sidewalk.

Existing hybrid systems models also focus on deterministic discrete transitions and have not considered multiple possible futures when predicting pedestrians behaviors [21, 77, 94]. Furthermore, existing hybrid system models primarily focus on laterally approaching pedestrians at midblocks with the intent to cross and do not consider other locations of pedestrian-vehicle interactions or pedestrians walking parallel to the road.

The current work is primarily focused on predicting the most probable pedestrian trajectory [23, 35, 77, 82, 85, 90]. However, future pedestrian behaviors are multimodal, i.e., they can take multiple possible paths, and their trajectories vary based on the actions taken. Several recent studies thus focused on multimodal predictions [26–28, 39, 86, 95–97]. A common approach is using conditional variational auto-encoders (CVAEs) [28, 86, 96, 98]. The CVAE approach samples trajectories based on a latent discrete variable representing different pedestrian behaviors such as "aggressive", "conservative", etc., or decisions such as "turn left/turn right". The sampling distributions are learned from the datasets using a deep learning framework. Another approach for multimodal prediction is to estimate Mixture Density Networks (MDNs) [99, 100], which parametrize a mixture of distributions using deep neural networks. Prokudin et al. [99] used MDNs to estimate the skeletal pose and, in

turn, the trajectories of pedestrians. Both CVAE and MDN approaches are computationally expensive. On the contrary, real-time AV operations require computationally efficient prediction models to allow the AVs to respond safely to pedestrians quickly. Moreover, the CVAE methods predict by sampling a latent variable which is not intuitive.

Each of the above approaches has certain limitations. A major limitation of the deep learning approaches is the requirement of large amounts of labeled training data. Though these models can generalize to both structured and unstructured scenes available in the training data, they tend to overfit the training data, making them less generalizable to new unseen environments. Moreover, the developed models are not interpretable, and the relevance of different features and model applicability to different scenarios is unknown. The data-driven deep learning models generally perform well in unstructured environments such as shared spaces. However, they do not explicitly account for the complex road behaviors, such as stop-and-go behaviors in structured urban scenarios.

Though the physics-based hybrid models require fewer data than deep-learning methods and are more generalizable to new environments, the current approaches have limited the predictions to short durations ($1-2\,s$) and limited pedestrian interactions (e.g., laterally crossing pedestrians). The discrete states of the road agents are usually maneuvers performed by the vehicle such as turn left, turn right, accelerate, and stop [92], or maneuvers performed by the pedestrian such as walk, run, and stop [21, 94]. However, it is more relevant for the AVs to predict the decision choices of the agents than their actual maneuvers, which these models do not entirely capture. For example, an AV might be more interested in knowing if a pedestrian walking on the sidewalk is going to cross or not, rather than if they are going to walk, run, or stop on the sidewalk. Moreover, existing hybrid systems models have focused on deterministic discrete transitions and have not considered multiple possible futures when predicting road agent behaviors.

To summarize, no existing pedestrian model produces trajectory predictions that satisfy all requirements, namely capturing multimodal pedestrian behaviors, predicting for a long term, not making overly conservative predictions, using computationally efficient algorithms, and not being overly data intensive. Moreover, current pedestrian models do not capture the complex stop-and-go behavior in structured urban environments. We address these gaps in Chapter IV.

## 2.4 Interaction modeling

The dynamic interactions between pedestrians greatly influence how their behaviors evolve over time [14, 28, 84]. Thus, modeling the interaction between pedestrians could improve the predictions of their behaviors and trajectories.

Existing models for multi-pedestrian trajectory predictions differ in how they incorporate the interactions with neighboring pedestrians. Traditionally, prediction models have modeled limited interactions such as the interactions between the ego-vehicle and the pedestrian [21, 101, 102]. These interactions have usually been modeled using Gaussian processes [76], Dynamic Bayesian Networks [21, 102], Social Force models [94], and velocity obstacles [103].

Current state-of-the-art deep learning methods implicitly model dynamically changing interactions between multiple pedestrians [28, 73, 75, 82, 83, 96]. Some studies segregate neighboring pedestrians into different zones based on their proximity to the pedestrian of interest. The hidden states abstracted from Long-Short Term Memory (LSTM) neural networks from the various zones are combined and used as model parameters [73, 82, 98] to predict the pedestrian trajectory, a technique called *pooling*. However, pooling hidden states of neighboring pedestrians lacks intuition making it challenging to interpret the predictions in a meaningful manner for AV planning. Other studies expressed the interaction between neighboring pedestrians as a spatiotemporal graph with nodes and edges. The nodes of the graph represent the road agents (pedestrians and ego-vehicle), and the edges represent their pairwise interactions that are based on the distance between the users [26, 28, 83, 86]. These methods have been successfully used in roadway interactions with multiple heterogeneous agents. Existing works [75, 96] have evaluated the performance of their models on real-world datasets [104, 105]. However, the interaction region usually is arbitrarily defined. Moreover, these approaches model the interactions' influence on the agents' continuous trajectory, which is computationally expensive.

To summarize, existing state-of-the-art pedestrian models lack intuition regarding the effects of other pedestrians on the predicted behaviors with arbitrarily defined interaction regions. We address these gaps in Chapter V.

## 2.5 AV planning

Automated driving in complex urban scenarios requires interactive and cooperative decision-making. Other agents' intentions need to be deduced and integrated

into a planning framework that allows for reasonable, cooperative decision-making without the need for inter-vehicle communication. Schwarting et al. [29] recently summarized existing motion planning and decision-making models for AVs. We focus on existing methods developed for cooperative and safe AV planning.

Typically, AV motion planners take in high-level planned routes and local environmental constraints (e.g., static obstacles, other vehicles, pedestrians) and output a trajectory for the low-level controller to track. AV motion planning has traditionally followed a pipeline approach and has been dealt with as a separate problem from predicting the road agents' behaviors with the assumption that the predictions over the trajectories of the road agents are known. For example, [106, 107] utilize predictions of pedestrian trajectories for AV planning where the planning stage is decoupled from the pedestrians' trajectory prediction. However, the predicted pedestrian trajectories may deviate from the actual behavior, and the planned AV actions may no longer be safe or optimal. To avoid potential collisions in such dynamic situations, AVs frequently replan their paths. Thus the rate of replanning must be high enough to compensate for the dynamically varying environment [31].

Recently, researchers have used hierarchical game theory for interactive AV planning [108–110]. The AV has a model of the pedestrian motion and performs nested optimization [110]. The AV assumes that the pedestrian's actions are optimized for an initial planned AV trajectory and predicts the pedestrian's trajectory. The AV then replans for the predicted trajectory of the pedestrian by maximizing its reward. Though this approach is promising, it has been applied to only discretized action and state space and a limited number of agents. These game theory approaches are computationally extensive and do not scale well to multiple road agents. Recently, some approaches have modeled the interaction between multiple agents and the AV using hierarchical game theory, which can potentially address the problem of scaling [111].

Existing motion planners incur a trade-off: they typically must attempt to encourage either safety and persistent feasibility or performance (quickly and successfully reaching the destination). Reachability-based approaches are a standard class of motion planning methods typically employed for guaranteeing safety [112–114]. These methods precompute a reachable set containing the robot's motion, then use these reachable sets to ensure collision avoidance at runtime. The precomputed reachable sets can be used to synthesize safe tracking controllers and incorporate uncertainty in the dynamics. However, these methods focus on real-time planning with guaranteed safety at the expense of some performance. These methods are also overly conserva-

tive as they consider all possible futures of the pedestrians through their reachable sets for collision avoidance.

To summarize, existing AV planning methods either plan for discrete action choices or use predictions from simple pedestrian models. We address the above gap in Chapter VI by developing a behavior-aware controller for the AV based on predictions from the developed pedestrian model.

# CHAPTER III

# Pedestrian Interaction with Automated Vehicles

## 3.1 Overview of chapter

AVs are a new and growing technology that are not yet widely available commercially. As a result, pedestrian interaction with AVs is limited and not clearly understood. A lack of understanding of pedestrian behavior around AVs can result in developing AV motion planning algorithms that negatively impact pedestrian trust and, in turn, acceptance of AVs. Pedestrian behavior, in large part, depends on them understanding the intentions of the road agents they are interacting with, including the AVs citerasouli2019autonomous, rudenko2020human. As discussed in Section 2.2, pedestrians have expressed concerns over not understanding AV intent [51, 52]. Uncertainty in AV intent, because of uncertain information from the AV or the surroundings, can result in critical situations [62]. Further, it is unclear how pedestrian behavior is different in the interactions with AVs compared to their interactions with HDVs.

This chapter presents the first main contribution of this dissertation, namely, the characterization of the role of driving behavior and vehicle type (HDV or AV) on pedestrian trust and pedestrian behaviors. Two user studies were performed. The findings of the first user study are published in [33, 34] and can be summarized as follows. In addition to confirming existing research that AV intent can be deduced from their behaviors [18], we found that pedestrian trust and behavior are influenced by AV driving behaviors. We found that pedestrians engaged in observable trusting behaviors when they reported trusting the AVs. However, the influence of AV driving behavior on pedestrian trust and behavior is valid only at unsignalized crosswalks. From the second user study, we found that pedestrians have different expectations towards AVs than HDVs, resulting in different pedestrian trust for AVs compared to HDVs. We also found that pedestrians' trust in the AVs increases with each

subsequent interaction. The second user study was a pilot, and the results were not published.

The first user study (N=30) examined pedestrian behavior under different levels of driving behavior and traffic signal types in a VR environment as discussed in [34]. The results of this study suggest that clarifying AV intent either through its driving behavior or through the traffic signals increases pedestrians' trust in AVs. Further, increased trust resulted in a significant increase in pedestrians' trusting behaviors, such as willingness to cross. Thus, trust in the AVs is found to directly influence pedestrian behavior.

The second user study (N=14) examined and compared pedestrian behavior around AVs and HDVs in a controlled, real-world environment. Results suggested that pedestrians were more willing to cross when the vehicles (both AVs and HDVs) drove less aggressively, similar to the relation between driving behavior and trust in AVs in the first user study. Interestingly, however, they thought human drivers were generally more aggressive than AVs, even though both had similar velocity profiles.

## 3.2 Pedestrian-AV interaction in virtual reality

In this section, we discuss the user study conducted in a VR environment — the reasoning behind hypotheses formation, the experimental setup used, the study methodology, the results obtained, and its implications. This user study examines the influence of information uncertainty on pedestrian trust in the AVs and, in turn, their behaviors.

*Uncertainty* is defined as the "inability to predict another's behavior because of a lack of information about the person or environment" [115, 116]. When individuals meet, they communicate and exchange information to reduce uncertainty with regards to each other's intentions. The more information gained, the less uncertainty one has about the other individual or situation. However, when direct communication with an individual is not possible, as in the case of pedestrian-AV interactions (see Section 2.2), people seek information from third parties or through observation [117].

As uncertainty increases, humans are more motivated to engage in information seeking to reduce uncertainty. Further, uncertainty decreases as the amount of information communicated increases. In other words, the more uncertainty, the more people seek information to reduce it; the more information provided, the less uncertainty. Trust, and uncertainty are inversely related [118, 119]. The greater uncertainty one has about the outcome of an interaction with an agent, the less trust one has

in that agent [118, 120]. Likewise, the more trust someone has in an agent, the less the uncertainty regarding the outcome of an interaction with that agent. *Trust* is defined as the "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [66]. In this chapter, trust is the attitude of the pedestrians that the AV would help them in their goal to cross the street.

Availability of information thereby plays an integral role in improving trust by reducing uncertainty. For example, [121] found that when the AV informed the human driver of the uncertainty in its ability to drive, the trust in the AV increased. In this chapter, we consider the information about AV's intent to be available from the AV's driving behavior and the traffic signal. Unlike [121], we do not quantify uncertainty but use the qualitative relationship between trust and uncertainty from AV driving behavior and/or absence of traffic signal to develop our hypotheses.

### 3.2.1 Hypotheses

In this study, we used uncertainty and availability of information to understand a pedestrian's trust in AVs. When a pedestrian approaches a crosswalk, there is some degree of uncertainty about an AV's actions – Will the AV stop? If so, will it stop within a safe distance, and will it remain stopped to allow the pedestrian to cross safely? Pedestrians attempt to reduce this uncertainty by seeking information to help them predict the AV's actions. The AV's actions can be directly estimated from the AV's driving behavior and/or can be obtained from the traffic signals that determine the right of way for all traffic participants. The more information available to facilitate the pedestrian's prediction of the AV's actions, the less uncertainty and the more trust they should have in the AV. Conversely, the less information available, the more uncertainty and the less trust they should have in the AV.

Thus the premise of this study is:

*As the available information that allows the pedestrian to predict the actions of the AV increases, so should trust in the AV.*

We considered two sources of information: the *AV's driving behavior* and the *traffic signal*. Driving behavior is typically classified into defensive, normal, and aggressive behaviors [122–124]. Defensive driving is characterized as slow and predictable, normal driving less so, while aggressive driving is characterized by unpredictable behavior including high speeds, delayed stopping or not yielding the right-of-way [122, 124]. For example, a defensively driving AV might sense a pedestrian trying to cross

the street and could start to slow down very early to indicate its intention of yielding to the pedestrian, even though legally, it may have been the AV's right-of-way. On the other hand, an AV driving more aggressively might slow down and yield late or even accelerate to indicate that it is not yielding to the pedestrian. Thus in scenarios where a pedestrian walks onto the road, an aggressive AV would brake later and harder than a defensive AV to avoid a potential collision with the pedestrian [64]. This makes it hard to predict whether an aggressive AV would ever slow down or stop for a pedestrian. The unpredictability of aggressive driving should lead to low trust in AVs. Thus, the more aggressive the vehicle drives, the more uncertain its behavior, and the lower the trust in AVs.

Pedestrians can also gather information from the surroundings – road type, stop sign location, traffic signal, etc. Vehicles are expected to stop at traffic signals. Thus the state of the signal would provide information about what the vehicles are expected to do. Particularly, AVs are expected to be more law-abiding, and thus, their intent would be more predictable. Therefore, signalized crosswalks should decrease uncertainty and increase trust in AVs by clarifying who should stop, whereas at unsignalized crosswalks, the right-of-way is less clear [32].

Furthermore, the crosswalk type should moderate the impacts of aggressive driving. The presence of a traffic signal should reduce the negative impact of aggressive driving on trust in AVs. Individuals should be more likely to believe that the AV will stop regardless of its driving behavior. Therefore, aggressive driving should have a weaker impact on trust in AVs at signalized crosswalks.

We also examine trusting behavior, which are actions that can be considered as descendants of trust [25]. Trusting behaviors are actions that increase the vulnerability of the pedestrians, such as crossing close to the vehicles, waiting for only a short duration before crossing, less gaze at oncoming vehicles, walking slowly on the road while crossing, etc. Trusting behaviors are related to trust but may also be moderated by workload, situational awareness, system capability, and other contextual and environmental factors [65–67]. Though the relation between trust and trusting behavior may not be straightforward, we still expect increased trust in the AVs to generally result in more trusting behaviors. Simply put, the more an individual trusts the AV, the more they should engage in trusting behaviors when interacting with AVs. Our research model is graphically summarized in Figure 3.1.

We test the following hypotheses:

***H1**: Aggressive AV driving behavior **decreases** pedestrians' trust in AVs.*

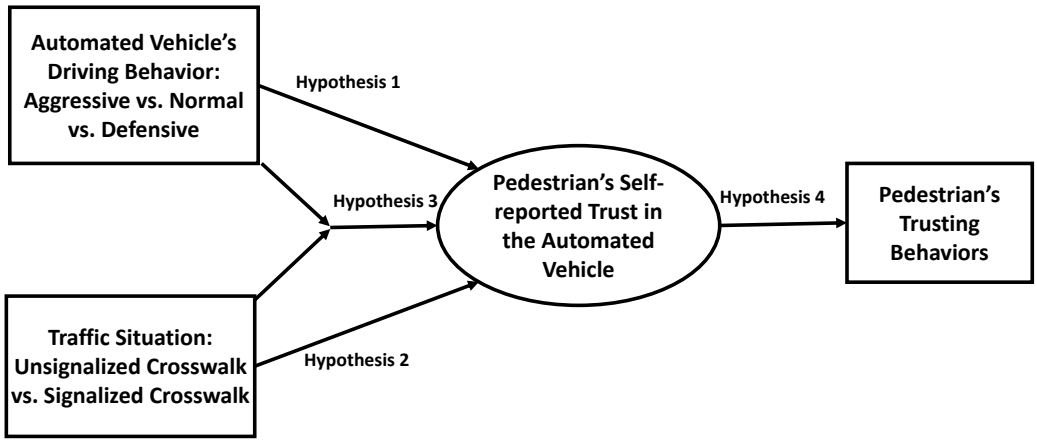



Figure 3.1: Pedestrian-AV Trust Model.

***H2**: Signalized crosswalk **increases** pedestrians' trust in AVs.*

***H3**: Crosswalk type **moderates** impact of aggressive AV driving behavior.*

***H4**: Pedestrians' trust in AVs **engenders more** trusting behaviors from pedestrians.*

### 3.2.2 Method

In this section, we describe the methodology employed in the user study to test the developed hypotheses. We conducted an experiment in a VR environment with simulated AVs. The details of the experiment are discussed in the following sections.

#### 3.2.2.1 Development of Experimental Setup

We developed a unique experimental setup to enable interactions between pedestrians and AVs in VR. Participants were placed in an immersive virtual environment (IVE) with an HTC Vive VR headset (Vive; HTC Corp., New Taipei, Taiwan), walking on an omni-directional treadmill (Virtuix Omni; Virtuix Inc., Austin, TX); they took on the role of a pedestrian walking in an urban environment. The left side of Figure 3.2 shows the equipment set up, while the right side shows the scene from participants' point-of-view as they wore the headset and walked on the treadmill. We developed the urban scenario simulation to be as realistic as possible. During the experiment, participants crossed a street at a mid-block crosswalk with several

oncoming AVs. The street was one-way with two lanes for the AVs. The AVs in this user study were fully automated (SAE Level 5 automation, see Figure 1.2) without any humans inside and produced engine sounds based on the speed of the AV and distance of the AV to the pedestrian. The IVE was built using the Unity Game Engine (Unity Technologies, San Francisco, CA).

The participant wears special shows fitted with IMU sensors and walks on the concave base of the treadmill. The ring surrounding the participant records their body orientation. The sensed feet movements and body orientation provide a direction and speed for movement in the IVE that matches the participant's input in the physical world. Unlike existing pedestrians simulators that had limited walking range [125–128], this simulator setup provides unlimited walking range to the pedestrians. This allows studying pedestrian-AV interactions not only when they are close to the road, but during the entire time they are in the environment. Additionally, the HTC Vive VR headset was fitted with a Pupil Labs eye tracker (Pupil; Pupil Labs, Berlin, Germany) to measure participant's gaze during the experiment.

We manipulated the type of driving behavior (defensive, normal, and aggressive) and the type of crosswalk (signalized and unsignalized). We employed a within-subjects experimental design, so every participant experienced all six conditions ($3 \times 2$). Sample videos of the six treatment conditions are available online[1] for reference.

#### 3.2.2.2 Experimental Task

In the experiment, participants were asked to move three numbered balls, one at a time, from one side of the street to the other, placing them in corresponding numbered boxes. Participants were required to remember the ball's number, which disappeared after it was picked up. The ball task was designed such that the crossing activity was embedded within the overall task of moving the balls. This served two purposes. First, it allowed participants to make multiple street crossings without experimenters explicitly instructing them to make such crossings. Second, the task was designed to reduce any participant reactivity, such as from an observer effect, wherein participants' actual crossing behaviors could be affected by their knowledge that the experimenters specifically measured such behaviors [129]. This ball task helped disguise from the participants that their crossing behaviors were of primary interest to the experimenters. The activities for moving a ball include approaching the crosswalk, waiting to cross, crossing the street, approaching the ball, picking up the ball, approaching the crosswalk, waiting to cross, crossing the street, approaching

[1] shorturl.at/lpAC4

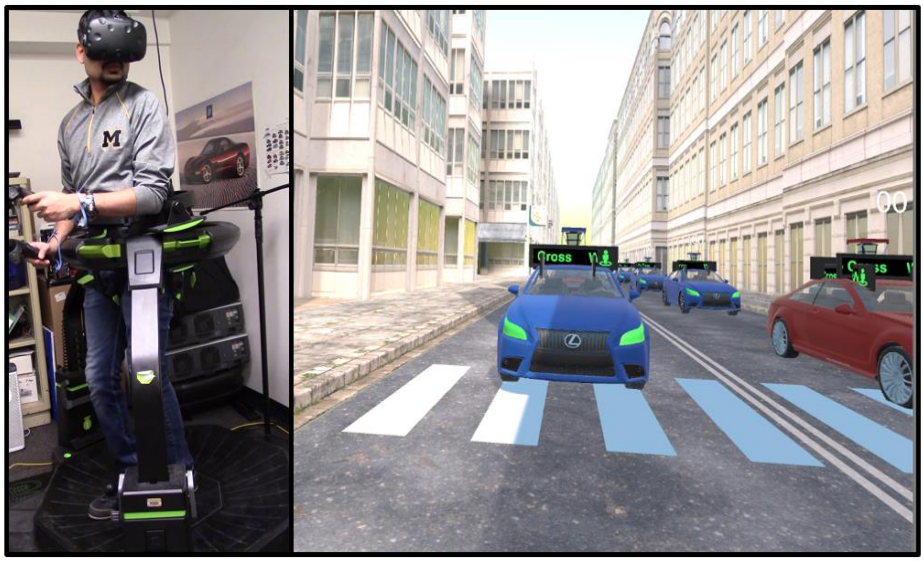


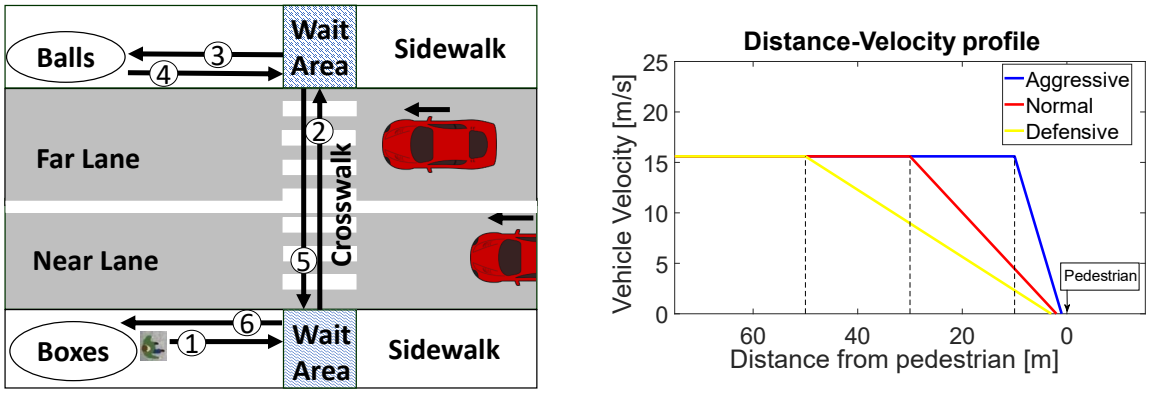

Figure 3.2: Virtual Reality setup for user study. The left side shows the user wearing the HTC Vive headset and walking on the omni-directional treadmill. The right side shows the virtual environment as seen by the participant.

the boxes, and dropping the ball. The numbers in Figure 3.3a (a) describe a typical sequence of pedestrian movements. Thus, by performing the ball task, they had to cross the street six times.

#### 3.2.2.3 Design of Interaction Scenarios

The simulation scenario was designed to mimic a downtown urban crosswalk. Figure 3.3a shows the layout of the VR environment. Participants could move around in all the different areas, including the sidewalks, the road lanes, and the wait areas. The wait areas are where the pedestrian would typically wait before crossing the road. Participants encountered AVs while crossing in either direction.

In both signalized and unsignalized conditions, all AVs approached the crosswalk at a constant speed of 15.6 m/s (35 mph). For each treatment condition, the vehicle changed its driving behavior when it encountered a pedestrian within its reaction distance (refer Table 3.1). This distance signifies the attentiveness of the different driving behaviors. As discussed earlier, the unpredictability of aggressive driving can be attributed to the delayed stopping or failing to yield the right-of-way [122] by the AV. We defined the aggressive driving behavior to be less cautious. We made AVs with aggressive driving behavior react to pedestrians much later than the defensive or normal driving conditions. We defined this behavior by varying the reaction distance,




(a) (b)

Figure 3.3: **(a)** Pedestrian state divisions in the virtual environment. Numbered arrows indicate a typical pedestrian path while doing the task. **(b)** Driving profiles for the three driving behaviors when the pedestrian is on the road in the same lane as the AV. To achieve the specified stopping distance and slow speeds, defensive behavior decelerated much earlier than normal or aggressive.

the distance from the pedestrian the AV would start reacting to pedestrians. We also varied the AVs' reactions to the pedestrian, such as stopping, slowing down, or not slowing based on the location of the AVs. The different driving behaviors were obtained by tuning the AVs' reactions, reaction distance, and driving parameters such as acceleration and speed across the three behaviors. The resulting behaviors were perceived to be different from one another during internal validation. The change in vehicle behavior is based on the discrete location of the pedestrian as described in Table 3.1. The stopping distance in Table 3.2 refers to the distance between the pedestrian and the vehicle when it is stopped.

Table 3.1: Different vehicle reactions to various pedestrian positions characterizing the different driving behaviors.

| **Behavior** | **Pedestrian Position** | | | | **Reaction dist.** |
|---|---|---|---|---|---|
| | **Sidewalk** | **Wait Area** | **Same lane** | **Other lane** | |
| Defensive | Full speed | Slow speed | Stop | Stop | 50 $m$ |
| Normal | Full speed | Slow speed | Stop | Slow speed | 30 $m$ |
| Aggressive | Full speed | Full speed | Stop | Full speed | 10 $m$ |

The cars always stop before the crosswalk if there is a pedestrian on the street. The cars do not stop when pedestrians are waiting/walking on the sidewalk. Even with the same driving behavior, the AVs do not react in the same way each time as their deceleration rates depend on the relative position between the pedestrian and AVs when the pedestrian reaches particular positions in the environment such as wait

Table 3.2: Vehicle parameters characterizing the different driving behaviors.

| **Behavior** | **Stopped distance** | **Maximum acceleration** | **Slow speed** | **Full speed** |
|---|---|---|---|---|
| Defensive | 3 $m$ | 3 $m/s^2$ | 4 $m/s$ | 15.6 $m/s$ |
| Normal | 2 $m$ | 5 $m/s^2$ | 7 $m/s$ | 15.6 $m/s$ |
| Aggressive | 1 $m$ | 8 $m/s^2$ | NA | 15.6 $m/s$ |

area, same lane, etc.

Table 3.1 provides the AV driving behavior model based on the pedestrian's positional state, and Table 3.2 provides the vehicle parameters used in the study. Typical driving profiles for the three driving behaviors are shown in Figure 3.3b. For a pedestrian in the wait area, as shown in Figure 3.3a, during the normal driving behavior conditions, the vehicles in the near lane and far lane slowed down from 15.6 m/s to 7 m/s. They started slowing down at a distance of 30 m from the pedestrian. When the pedestrian stepped into the near lane, the vehicle in the near lane stopped 2 m from the pedestrian, whereas the vehicle in the far lane continued at a speed of 7 m/s.

Additionally, in the signalized conditions, the AV stopped at the appropriate stopping distances (Table 3.2) when the signal was red or yellow and maintained the same behavior as in unsignalized conditions (Table 3.1) when the signal was green. In signalized conditions, when the pedestrian was not on the road, the stopping distance refers to the distance between the vehicle when it was stopped and the center of the crosswalk. The behaviors across signalized and unsignalized conditions were maintained to be as similar as possible for experimental validity and to avoid any confounding effects due to variations in the vehicle behaviors when examining the effect of the traffic signal.

The signal for the vehicles operated on a 38-second cycle: green for 20 seconds, yellow for 3 seconds, then red for 15 seconds [130]. The cycle ran continuously in the background, but the signal changed to yellow and red only after the participant pressed the provided signal button. If the participant did not press the signal button, the signal remained green. Vehicular traffic was generated in a predetermined pseudorandom sequence of short (3-second) and long (5-second) gaps. The probability of a short gap occurring was 75%, and a long gap was 25%, inducing the participants to observe the cars during the short gaps while waiting for a long gap to occur to cross. Vehicles were generated in both lanes of the street, going in the same direction.

#### 3.2.2.4 Experiment Design

We employed a within-subjects experimental design. Participants underwent a training session before testing to familiarize themselves with the equipment and the task. After the training session, participants experienced each of the six treatment conditions (defensive, normal, and aggressive driving behaviors for signalized and unsignalized crosswalks) once. The conditions' sequence was counterbalanced using a Latin square design [131]. The standard Latin square design we employed in the study is available in Appendix A.1.

The balanced Latin square design has a group of sequences of treatment conditions such that every condition appears before and after every other condition exactly once. This design helps to compensate for immediate sequential effects [131].

#### 3.2.2.5 Measurements

We collected attitudinal, behavioral, physiological, and other self-reported measures. We measured participants' propensity to trust [132] and experience with virtual reality before the experiment (calculated as a mean of 1-7 Likert scale responses). After each treatment condition, participants gave 1-7 Likert scale ratings measuring trust and perceived AV aggressiveness. For measuring self-reported trust, we adapted the Muir scale questions [133], a highly validated trust in automation scale. We modified the questions to reflect the pedestrian-AV interaction context (refer Appendix A.2). Self-reported trust was calculated as the mean of the responses to the trust questionnaire. We also measured simulator sickness, calculated using the items and procedure mentioned in [134] (refer Appendix A.3), at the end of the experiment.

We collected six dependent measures of trusting behavior from the simulation, some of which were calculated for each of the six crossings within a treatment condition and averaged. *Average distance to collision* measured how close a participant was to being hit by the AV as the distance between the AV in its lane and the participant when he/she entered that lane.

*Average jaywalking time* was the average time participants spent either crossing the street when the AV had the right-of-way, which was whenever the pedestrian signal was red in the unsignalized conditions or crossed the street outside of the crosswalk in both the signalized and unsignalized conditions. *Average wait time* measured the average time they spent waiting before they crossed the street. *Average crossing speed* measured how fast they crossed the street. *Average crossing time* was the average duration of the crossing. *Overall task time* measured how long they took to complete

the entire treatment condition.

We examined participants' eye gaze to explore its relationship with self-reported trust in AVs. A lack of monitoring is related to high trust in automation [135]. We divided the environment into seven areas of interest (AOI): (1) looking at approaching AVs, (2) checking for AVs, (3) pedestrian signal light, (4) traffic light, (5) task materials, (6) crosswalk and buildings directly across the crosswalk, and (7) everything else in the environment that included the sky, other buildings along the road, and roads not in the crosswalk region. The crosswalk and buildings directly across represented regions when a participant stared ahead. We measured the duration of time each participant spent looking at the different AOI using the Pupil Labs eye tracker.

In our user study, the AOI at which the participants gazed was identified in real-time by interfacing the Pupil Labs eye tracker with the Unity simulation. The Unity simulation obtained the gaze point from the eye tracker and cast a ray to identify which AOI intersected with the gaze ray at every sampling instant.

#### 3.2.2.6 Study Participants

We recruited participants through email and obtained informed consent from each participant. Thirty participants, of which 28 were college students, joined in this user study (9 female), with a mean age of 22.5 years (standard deviation [SD] = 2.8 years). The study population was relatively young as it appealed primarily to the student population in the university, and we did not explicitly control for age during subject recruitment.

This research complied with the American Psychological Association Code of Ethics and was approved by the institutional review board at the University of Michigan.

### 3.2.3 Results and discussion

The descriptives (mean and standard deviation) of our survey measurements are reported in Table 3.4. Our study had a within-subjects experimental design and collected repeated measurements (for each of the six treatment conditions) from the same subject. To account for this non-independence in the data, we employed mixed linear repeated modeling (MLRM) technique [136] to understand the relationships between the dependent and independent variables. MLRM makes it easy to study the effects of covariates and the treatment variables on the dependent variable. We

used SPSS v24 (IBM, Armonk, NY) for all our analyses.

#### 3.2.3.1 Manipulation Check of Aggressive Driving

We conducted a manipulation check to verify if the participants perceived each driving condition to have different levels of driving aggression. To accomplish this, we ran an MLRM with the driving condition as the independent variable and the perceived AV driving aggression as the dependent variable. The model revealed a significant difference ($p < 0.001$) among the driving conditions. The mean (standard deviation) values were $\overline{x} = 2.67$ (0.22) for defensive, $\overline{x} = 3.44$ (0.21) for normal and $\overline{x} = 4.24$ (0.23) for aggressive driving conditions. As shown in Figure 3.4, all pairwise comparisons were significantly different from one another ($p < 0.05$). Our results indicate that our manipulation of driving behavior was successful.

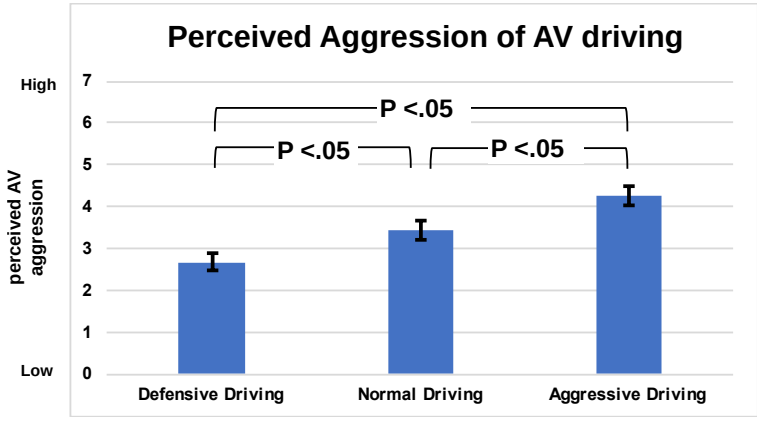



Figure 3.4: Manipulation Check of Aggressive Driving. Perceived Aggression of AV driving is lowest for Defensive Driving and highest for Aggressive Driving conditions.

#### 3.2.3.2 Measurement Validity

To verify if our survey constructs measured what they were intended to measure, we conducted a factor analysis to examine convergent and discriminant validity of the self-reported trust and simulator sickness measures (see Table 3.3). Only one item (Trust: Reliability) did not meet the 0.7 loading requirement indicating convergence validity. Also, no cross-loadings exceeded 0.3, indicating discriminant validity. Thus, the results indicate both discriminant and convergent validity [137].

Before conducting our analyses, we checked for heteroscedasticity by performing Glejser test, which states that variables have non-linear and unequal variances if $p$

$< 0.05$ [138]. We found evidence of both nonlinearity and unequal variances related to average distance to collision ($p = 0.01$), average jaywalking time ($p = 0.03$), and average crossing speed ($p = 0.01$). To improve the linearity and equality of the variances, we performed log transformations on each dependent variable and verified the absence of heteroscedasticity ($p \geq 0.05$ for all dependent variables).

Table 3.3: Survey measurement validity using factor and cross loadings of trust and simulator sickness survey measures

| Item | Self-reported trust in AVs | Simulator sickness (SS) |
|---|---|---|
| Trust: Competence | **0.83** | 0.12 |
| Trust: Predictability | **0.86** | 0.01 |
| Trust: Dependability | **0.86** | 0.09 |
| Trust: Responsibility | **0.84** | 0.05 |
| Trust: Reliability | **0.62** | 0.02 |
| Trust: Faith | **0.72** | 0.00 |
| SS: Disorientation | 0.07 | **0.91** |
| SS: Nausea | 0.06 | **0.82** |
| SS: Oculomotor | 0.01 | **0.91** |
| Convergent validity: factor loadings > 0.7; Discriminant validity: cross loadings < 0.3 | | |
| Bold values indicate the representative factors included in the measures in each column. | | |

#### 3.2.3.3 Population effects

We found that neither age (fixed effects estimate, $\beta = 0.03$, $p = 0.77$) nor gender (fixed effects estimate, $\beta =$ -0.22, $p = 0.40$) had significant effects on the self-reported trust in AVs. The effect of age was not significant perhaps because of the limited age range of our study population. The study had a fairly young population (18-30 years) with a mean age of 22.5 years (and standard deviation [SD] = 2.8 years).

#### 3.2.3.4 Hypothesis Testing

For our analysis, we used the self-reported perceived AV aggression as an independent variable because it is a more accurate measure of how the pedestrians perceived the AVs' behavior, which would affect their trust. We used the self-reported trust, calculated as the mean of the responses to the trust survey, as the dependent variable. We analyzed H1–H3 in two parts. First, we analyzed the main effects model with the control variables and the variables measuring aggressive driving (self-reported perceived AV aggression) and crosswalk type. Second, we included the moderation effect involving signalized crosswalks and aggressive driving. We employed the full model with the moderation effect because it had a lower Schwarz's Bayesian Criterion (BIC

= 472) than the model with only the main effects (BIC = 500) and thus fit the data better [139]. The full model and correlations are shown in Tables 3.4 and 3.5.

Table 3.4: Descriptives of measurements and correlations between the measurements.

| | Parameters | Mean | S.D. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Trust | 5.68 | 1.10 | | | | | | | |
| 2 | Aggressive driving | 3.47 | 1.83 | **-0.47**** | | | | | | |
| 3 | Signalized crosswalks | 0.50 | 0.50 | **0.33**** | **-0.31**** | | | | | |
| 4 | Driving condition | 0.50 | 0.50 | -0.08 | **0.28**** | 0.00 | | | | |
| 5 | Age | 22.50 | 2.76 | 0.06 | -0.06 | 0.00 | 0.00 | | | |
| 6 | Propensity to trust | 5.33 | 0.46 | **0.20**** | **-0.31**** | 0.00 | 0.00 | -0.14 | | |
| 7 | Virtual reality experience | 3.36 | 1.20 | 0.01 | -0.01 | 0.00 | 0.00 | **-0.34**** | **0.20**** | |
| 8 | Simulator sickness | 28.30 | 23 | -0.10 | **0.31**** | 0.00 | 0.00 | **0.22**** | **-0.42**** | **-0.24**** |
| ** Correlation is significant at the 0.01 level (2-tailed). | | | | | | | | | | |
| **Trust** is calculated as the mean of the responses (1-7 Likert scale) from the survey in Apppendix A.2<br>**Aggressive driving** is a rating of perceived aggression (1-7 Likert scale) of the AV driving.<br>**Signalized crosswalks** is a Boolean variable for presence/absence of traffic signal.<br>**Driving condition** is a Boolean variable for low-aggressive and high-aggressive behavior.<br>**Propensity to trust** is measured on a 1-7 Likert scale from the Complaceny rating survey [132].<br>**Virtual Reality experience** is measured on a 1-7 Likert scale before the experiment starts.<br>**Simulator sickness** is calculated using the formula in [134]. | | | | | | | | | | |

We derived our mixed linear model from both level 1 (equation 3.1) and level 2 (equation 3.2) equations [140]. This two-level modeling accounts for the effects of both group-level and individual-level variables and allows random variations for the group-level variables [136]. In the level 1 equation, $Y_{ij}$ is the trust outcome for individual $i$ (from 30 subjects) in group $j$ (from 6 treatment conditions). $\beta_{0j}$ represents the group intercept values, $\beta_{1j}$ and $\beta_{2j}$ represent the effects of group predictors $SignalCond_j$ and $DriveCond_j$ respectively, whereas $\beta_{01}$, $\beta_{02}$, $\beta_{03}$, $\beta_{04}$, and $\beta_{05}$ represent the effects of the individual predictors $Aggress_{ij}$, $Age_{ij}$, $ProTrust_{ij}$, $VirReaExp_{ij}$, and $SimSic_{ij}$, respectively. $\epsilon_{ij}$ represents the residual for individual $i$ in group $j$.

$$Y_{ij} = \beta_{0j} + \beta_{1j}(SignalCond_j) + \beta_{2j}(DriveCond_j) + \beta_{01}(Aggress_{ij}) + \beta_{02}(Age_{ij}) + \beta_{03}(ProTrust_{ij}) + \beta_{04}(VirReaExp_{ij}) + \beta_{05}(SimSic_{ij}) + \epsilon_{ij} \tag{3.1}$$

Group level variables are associated with varying intercepts shown in the level 2 equations (equation 3.2). Gammas $\gamma_{00}$, $\gamma_{10}$, $\gamma_{20}$ represent the intercepts (fixed main effects) while $\nu_{0j}$, $\nu_{1j}$, and $\nu_{2j}$ represent their corresponding variances. These variances highlight that $\beta_{0j}$, $\beta_{1j}$ and $\beta_{2j}$ are allowed to randomly vary. Gammas $\gamma_{01}$, $\gamma_{02}$, $\gamma_{03}$, $\gamma_{04}$, and $\gamma_{05}$ represent the intercepts (fixed main effects) for their corresponding individual-level counterparts, whereas $\gamma_{11}$ represents the effect of the interaction term $Aggress_{ij} * SignalCond_j$. $\beta_{01}$, $\beta_{02}$, $\beta_{03}$, $\beta_{04}$, and $\beta_{05}$ are not allowed to randomly vary

Table 3.5: Trust model: Higher trust during less aggressive driving and during presence of signal with presence of signal moderating the effect of aggressive driving on trust.

| Independent parameter | | Estimation ($\beta$) | S.E. | df | t | Sig. | 95% C.I. | |
|---|---|---|---|---|---|---|---|---|
| **Intercept*** | ($\gamma_{00}$) | 4.93 | 0.23 | 63.76 | 21.86 | 0.00 | 4.49 | 5.39 |
| **Aggressive driving*** | ($\gamma_{01}$) | -1.08 | 0.22 | 76.96 | -4.97 | 0.00 | -1.51 | -0.65 |
| **Signal condition*** | ($\gamma_{10}$) | 0.41 | 0.12 | 85.16 | 3.50 | 0.00 | 0.18 | 0.65 |
| **Aggressive driving × signal condition*** | ($\gamma_{11}$) | 0.40 | 0.12 | 91.67 | 3.43 | 0.00 | 0.17 | 0.64 |
| Driving condition | ($\gamma_{20}$) | 0.10 | 0.06 | 49.68 | 1.76 | 0.08 | -0.01 | 0.21 |
| Age | ($\gamma_{02}$) | -0.01 | 0.12 | 21.47 | -0.08 | 0.93 | -0.27 | 0.25 |
| Propensity to trust | ($\gamma_{03}$) | 0.11 | 0.13 | 21.41 | 0.88 | 0.39 | -0.15 | 0.38 |
| Virtual reality experience | ($\gamma_{04}$) | -0.01 | 0.12 | 21.286 | -0.03 | 0.97 | -0.26 | 0.25 |
| Simulator sickness | ($\gamma_{05}$) | 0.03 | 0.13 | 22.72 | 0.21 | 0.83 | -0.24 | 0.30 |
| | | | | | | | | |
| Independent parameter | | Estimation ($\beta$) | S.E. | Wald Z | | Sig. | 95% C.I. | |
| Random intercept variances | ($\nu_{0j}$) | 0.21 | 0.14 | 1.49 | | 0.14 | 0.06 | 0.77 |
| Random signal condition variances | ($\nu_{0j}$) | 0.03 | 0.03 | 0.99 | | 0.32 | 0.00 | 0.24 |
| Random driving condition variances | ($\nu_{0j}$) | 0.01 | 0.02 | 0.64 | | 0.52 | 0.00 | 0.21 |
| Residual variances | ($\epsilon_{ij}$) | | | | | | | |
| **Treatment condition 1 (Defensive Unsignalized)*** | | 0.98 | 0.28 | 3.52 | | 0.00 | 0.56 | 1.71 |
| **Treatment condition 2 (Normal Unsignalized)*** | | 0.52 | 0.16 | 3.35 | | 0.00 | 0.29 | 0.94 |
| **Treatment condition 3 (Aggressive Unsignalized)*** | | 1.14 | 0.33 | 3.44 | | 0.00 | 0.65 | 2.02 |
| **Treatment condition 4 (Defensive Signalized)*** | | 0.21 | 0.08 | 2.77 | | 0.01 | 0.11 | 0.44 |
| **Treatment condition 5 (Normal Signalized)*** | | 0.31 | 0.09 | 3.38 | | 0.00 | 0.01 | 0.44 |
| **Treatment condition 6 (Aggressive Signalized)** | | 0.07 | 0.06 | 1.10 | | 0.27 | 0.01 | 0.39 |
| *Significant model parameters. | | | | | | | | |
| Fixed effects estimates ($\beta$) indicate the direction and degree of relationship between trust and the model variables. | | | | | | | | |

and therefore do not have corresponding variances.

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \nu_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(Aggress_{ij}) + \nu_{1j} \\
\beta_{2j} &= \gamma_{20} + \nu_{2j} \\
\beta_{01} &= \gamma_{01} \\
\beta_{02} &= \gamma_{02} \\
\beta_{03} &= \gamma_{03} \\
\beta_{04} &= \gamma_{04} \\
\beta_{05} &= \gamma_{05}
\end{aligned} \tag{3.2}$$

Our mixed linear model is derived by substituting equation 3.2 into equation 3.1. The final model we used is shown in equation 3.3.

$$Y_{ij} = \gamma_{00} + \gamma_{01}(Aggress_{ij}) + \gamma_{02}(Age_{ij}) + \gamma_{03}(ProTrust_{ij}) + \gamma_{04}(VirReaExp_{ij}) + \gamma_{05}(SimSic_{ij}) + \gamma_{10}(SignalCond_j) + \gamma_{11}(Aggress_{ij}) * (SignalCond_j) + \gamma_{20}(DriveCond_j) + \nu_{0j} + \nu_{1j}(SignalCond_j) + \nu_{2j}(DriveCond_j) + \epsilon_{ij} \tag{3.3}$$

We also tested H4 using a mixed linear modeling approach (Table 3.6). We used the mean of trust in the AV per condition as the independent variable when predicting trusting behaviors. To justify the aggregation by condition, we calculated the intraclass correlation coefficient (ICC). ICC measures the degree to which an individual-level variable is influenced by group-level membership. The higher the ICC, the more the individual-level variable is driven by group membership, and the more justification one has to create a group-level variable. The ICC value of trust in the AV per condition was 0.44, exceeding 0.10 [141], indicating that a group variable was valid.

H1 posited that aggressive driving decreases trust in the AV; this was supported (fixed effects estimate, $\beta$ = -0.17, $p < 0.001$). H2, which stated that signalized crosswalk increases trust in the AV, was supported (fixed effects estimate, $\beta = 0.53$, $p < 0.001$). Figure 3.5 shows the main effect of the signalized crosswalk vs. unsignalized crosswalk on trust in the AV. Finally, H3 was examined in the full model (Table 3.5). H3, the impact of aggressive driving on trust depends on the type of crosswalk, was also supported (fixed effects estimate, $\beta = 0.38$, $p < 0.001$) shown in Figure 3.6.

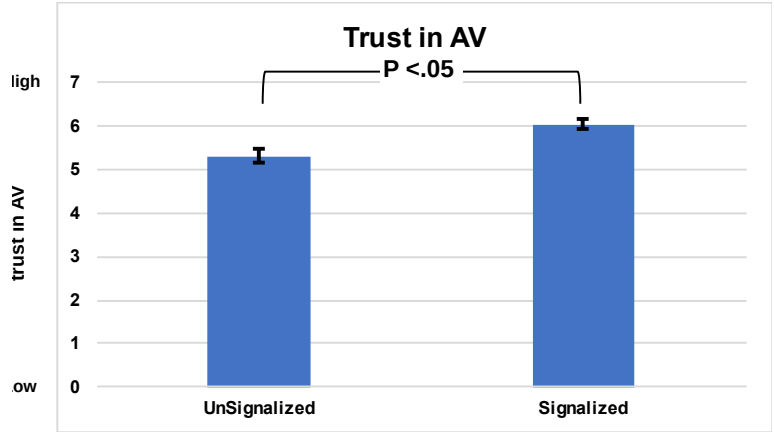



Figure 3.5: Main Effects of Signalized Crosswalks. Higher self-reported trust in AVs in Signalized conditions.

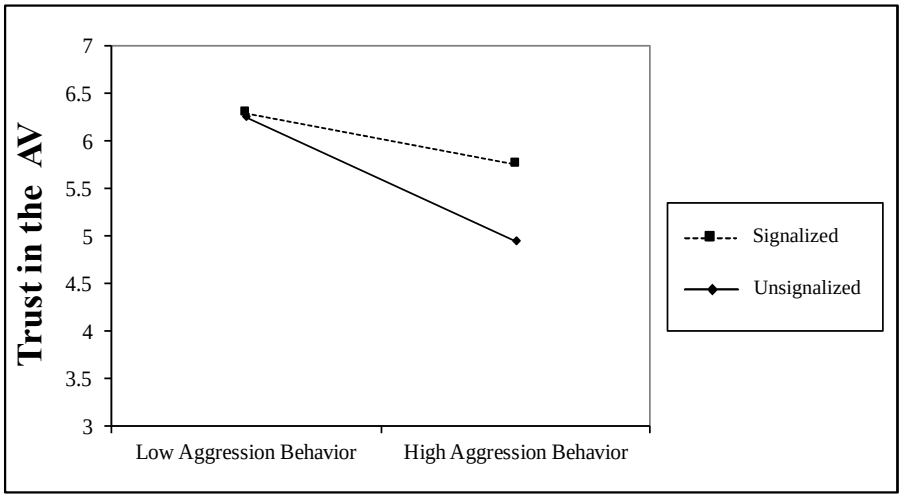



Figure 3.6: Moderation of Aggressive Driving by Signalized Crosswalks. Trust reduction due to high aggression behavior is lower for Signalized than Unsignalized Crosswalks.

H4, which stated that trust in the AV leads to more trusting behaviors, was partially supported. We defined trusting behavior as behavior that prolongs a participant's exposure to being vulnerable to the AV's actions. Therefore, when participants trusted the AV, we expected participants to cross closer to AVs, resulting in decreased average distance to collision, crossing earlier, decreasing wait time and overall task time, and walking slowly, resulting in decreased crossing speed. We also expected the participants to take more risks and cross when it was not their right-of-way resulting in increased jaywalking time and increased crossing time due to decreased crossing speed. We employed an MLRM for each of these objective trust behavior measures, with the objective measure being the dependent variable and self-reported trust being the independent variable.

Trust in the AV was significantly related to average distance to collision (fixed effects estimate, $\beta$ = -0.38, $p < 0.001$), average jaywalking time (fixed effects estimate, $\beta = 0.17$, $p < 0.05$), average waiting time (fixed effects estimate, $\beta = 0.18$, $p < 0.001$) and overall task time (fixed effects estimate, $\beta = 32.31$, $p < 0.001$). In other words, the more participants trusted the AV, the closer they came to the AV while crossing, the more they jaywalked, the longer they waited to cross and more time it took for them to complete the task. Trust in the AV was not related to average crossing time (fixed effects estimate, $\beta$ = -0.08, $p > 0.05$) or average crossing speed (fixed effects estimate, $\beta = 0.02$, $p > 0.05$; see Table 3.6).

Table 3.6: Mixed linear models of trust and each trusting behavior separately with trust being the dependent variable predicting trusting behaviors.

| **Parameter** | **Estimation ($\beta$)** | **S.E.** | **df** | **t** | **Sig.** | **95% C.I.** | |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| Intercept | 0.88 | 0.01 | 75.63 | 87.43 | 0.00 | 0.86 | 0.90 |
| C.L.Trust | -0.38 | 0.01 | 68.60 | -38.71 | 0.00 | -0.40 | -0.36 |
| Dependent Variable: **Log of average distance to collision*** (m) | | | | | | | |
| | | | | | | | |
| Intercept | -0.08 | 0.04 | 17.37 | -1.79 | 0.09 | -0.17 | 0.01 |
| C.L.Trust | 0.17 | 0.04 | 13.84 | 3.78 | 0.00 | 0.07 | 0.26 |
| Dependent Variable: **Log of average jaywalking time*** (s) | | | | | | | |
| | | | | | | | |
| Intercept | 16.24 | 0.53 | 92.74 | 30.60 | 0.00 | 15.19 | 17.30 |
| C.L.Trust | 4.26 | 0.54 | 63.37 | 7.92 | 0.00 | 3.19 | 5.34 |
| Dependent Variable: **Average waiting time*** (s) | | | | | | | |
| | | | | | | | |
| Intercept | 222.58 | 4.33 | 132.34 | 51.45 | 0.00 | 214.02 | 231.14 |
| C.L.Trust | 32.31 | 4.24 | 75.75 | 7.63 | 0.00 | 23.88 | 40.76 |
| Dependent Variable: **Overall task time*** (s) | | | | | | | |
| | | | | | | | |
| Intercept | 3.86 | 0.10 | 165.42 | 39.40 | 0.00 | 3.67 | 4.06 |
| C.L.Trust | -0.08 | 0.10 | 105.36 | -0.82 | 0.42 | -0.28 | 0.12 |
| Dependent Variable: Average crossing time (s) | | | | | | | |
| | | | | | | | |
| Intercept | 0.27 | 0.01 | 173.80 | 32.28 | 0.00 | 0.25 | 0.29 |
| C.L.Trust | 0.01 | 0.01 | 108.91 | 0.70 | 0.49 | -0.01 | 0.02 |
| Dependent Variable: Log of average crossing speed (m/s) | | | | | | | |
| | | | | | | | |
| **C.L. Trust** = Condition Level Trust, mean of trust for each treatment condition. All six **trusting behaviors** are measured from the simulation. | | | | | | | |
| Fixed effects estimates ($\beta$) of the models indicate the direction and degree of relationship between trust and trusting behaviors. | | | | | | | |
| *Behaviors with significant relationship with trust. | | | | | | | |

Following prior literature, a lack of visual monitoring of the automation can be viewed as an act of trusting behavior [135]. Therefore, we expected that trust in AVs would negatively correlate with gaze at AVs. To better understand the relationship between monitoring and trust, we divided our analysis by one of three actions: waiting, crossing, and tasking. *Waiting* included the time a participant spent waiting to cross the street. *Crossing* included the actual walking across the street. *Tasking* included the remaining time spent working on the task of moving the balls. We calculate gaze ratios per action by dividing the duration a participant focused on a particular area by the action type's total duration (Table 3.7). Then we conducted a repeated measure correlation between each gaze area ratio and trust in the AV (Table 3.8).

Table 3.7: Gaze distribution by AOI and driving behavior condition.

| **Areas of Interest (AOI)** | **Defensive** | **Normal** | **Aggressive** | **Overall** |
|---|---|---|---|---|
| AVs approaching the crosswalk | 25.0% | 18.1% | 13.4% | 18.7% |
| Checking for AVs (Looking in the general direction of AVs when no AVs are present on the road) | 2.0% | 2.3% | 4.4% | 2.9% |
| Crosswalk and buildings across the crosswalk | 35.9% | 38.0% | 40.6% | 38.4% |
| Task elements (racks on either side of crosswalk) | 11.5% | 12.2% | 13.2% | 12.3% |
| Pedestrian signal light on either side of crosswalk | 3.4% | 3.5% | 3.7% | 3.6% |
| Traffic light | 0.7% | 0.7% | 0.4% | 0.6% |
| All other areas | 21.5% | 25.2% | 24.3% | 23.5% |

Trust in the AV was negatively related to monitoring. Time spent looking at the approaching AVs while crossing was negatively correlated with trust in the AV for normal and aggressive driving. Also, there was a negative correlation between self-reported trust in AVs and time spent checking for AVs while crossing in normal driving behavior. Looking at the pedestrian light while crossing in normal behavior and looking at the traffic light while waiting and tasking in defensive behavior were negatively correlated with self-reported trust in AVs. These results support previous research suggesting decreased monitoring is related to increased trust [135]. While waiting at the crosswalk, gaze at the crosswalk and the buildings across the crosswalk indicate that the pedestrians were staring ahead and not monitoring the AVs. This time they spent looking at the crosswalk and buildings positively correlated with trust.

To summarize, the results are organized around four overarching findings. First, aggressive driving reduced pedestrian's self-reported trust in AVs. Second, signalized crosswalks increased pedestrian's self-reported trust in AVs. Third, the impact of driving behavior on pedestrian's self-reported trust in AVs varied significantly by the type of crosswalk. Finally, trust in AVs led to increases in some trusting behaviors, but not all.

In this study, we proposed hypotheses for the development of trust based on information availability. When two agents interact, the more information gained about the other agent, the less uncertain one is about the other agent. We highlight the importance of AV driving behavior and traffic signal and the moderation effect of a traffic signal on the impact of aggressive driving on pedestrians' trust in AVs.

Table 3.8: Repeated measures correlation between gaze and trust separated by activity and driving condition.

| | **Waiting** | | | **Crossing** | | |
|---|---|---|---|---|---|---|
| **Areas of Interest** | Defensive | Normal | Aggressive | Defensive | Normal | Aggressive |
| AVs approaching the crosswalk | -0.07 | -0.24 | -0.13 | -0.23 | **-0.34*** | **-0.25*** |
| Checking for AVs (looking in the general direction of AVs when no AVs are present on the road) | -0.22 | -0.11 | -0.13 | -0..16 | **-0.34*** | -0.24 |
| Crosswalk and buildings across the crosswalk | 0.24* | **0.42***** | **0.27**** | 0.13 | **0.38**** | 0.20 |
| Task elements (racks on either side of crosswalk) | 0.01 | -0.04 | -0.12 | 0.07 | 0.05 | -0.12 |
| Pedestrian signal light on either side of crosswalk† | -0.02 | -0.03 | 0.02 | -0.13 | -0.22* | 0.05 |
| Traffic light† | **-0.20*** | -0.14 | -0.02 | -0.02 | 0.04 | -0.07 |
| All other areas | -0.15 | **-0.32*** | **-0.31*** | -0.10 | **-0.37**** | -0.19 |
| † Pedestrian signal light and traffic light AOI available only during the three signalized conditions | | | | | | |
| A mixed linear model fitted between trust and each gaze ratio to calculate the repeated measures correlation. | | | | | | |
| * Correlation is significant at the 0.05 level (2-tailed). | | | | | | |
| ** Correlation is significant at the 0.01 level (2-tailed). | | | | | | |
| *** Correlation is significant at the 0.001 level (2-tailed). | | | | | | |

Specifically, we found that both sources of information, AV driving behavior and traffic signal, predicted pedestrians' trust in the AVs.

We systematically examined AV driving behavior and found that aggressive AV driving behavior significantly decreased trust in AVs. Thus driving behavior could implicitly convey the AV intent to pedestrians. This finding aligns with existing research that has found pedestrians generally prefer a conservative AV driving behavior [18, 19, 58].

Our work also calls attention to the importance of the presence of traffic signals in pedestrian-AV interactions. To the author's knowledge, the impact of traffic signals on pedestrian trust in AVs has not been explored before. We found that pedestrians, in general, trusted the signalized crosswalks more than the unsignalized crosswalks. This is in line with existing research in pedestrian-HDV interactions, which have reported increased trusting behavior such as lower crossing speeds, reduced gaze at vehicles, and shorter distances to collision at signalized crosswalks [142–144].

More importantly, we found that the influence of the AV's driving behavior is primarily determined by whether the crosswalk is signalized or unsignalized. Signalized crosswalks significantly reduced the negative effects of aggressive driving on trust. It could be because signalized crosswalks dictate the right-of-way, and AVs are expected to follow the right-of-way [60]. Thus, irrespective of their driving behavior, the AVs are always expected to stop when the pedestrian has the right-of-way. In any case,

our findings demonstrate the importance of incorporating the presence of traffic signal when understanding trust in the AV and help identify generalized situations during which pedestrians trust AVs. For example, trust is generally high in signalized conditions irrespective of the driving behavior (refer Figure 3.6). We highlighted the link between trust in the AV and trusting behaviors. When pedestrians under-trusted the AVs, they exhibited behaviors such as high distance to collision, fewer instances of jaywalking, increased looking at AVs while crossing, etc. As trust in the AV increased, pedestrians were much more willing to be vulnerable to the actions of the AV, which came in the form of reductions in distance to collision and increases in jaywalking. We also observed trusting behavior in the form of a lack of monitoring the AVs (i.e., low gaze ratio at the AVs when the self-reported trust scores were high), which aligns with existing research on drivers' trust in AVs [145]. This suggests that pedestrian behavior can vary based on their level of trust in the AV. Thus, it is necessary to maintain an appropriate level of trust for safe pedestrian-AV interactions.

## 3.3 Pedestrian-AV interaction in the real-world

The developed VR platform (Section 3.2.2.1) enables studying pedestrian interaction with AVs. However, it does not have the risks and goals of the real world and can thus result in different pedestrian behaviors compared to the real world [146]. As discussed in Section 2.2.3, it is still unclear how pedestrians will behave around AVs in the real world. To understand real-world pedestrian-AV interactions, we conducted a small-scale user study (N=14) at the Mcity test track located in Ann Arbor. We discuss the rationale behind hypotheses development, the experimental setup used, the study methodology, results obtained, and implications. This user study examines the influence of vehicle type on pedestrian trust in AVs.

### 3.3.1 Hypotheses

Surveys on public perception of AVs suggest that people are still skeptical of accepting AVs, and more than half the respondents stated they would not use an AV [147]. Pedestrians do not have any direct benefits, unlike the drivers/passengers of the AV [148]. Moreover, as discussed in Section 2.2, pedestrians can be warier of trusting AVs than HDVs. Thus we hypothesized that pedestrians would trust HDVs more than AVs.

This untrusting attitude of pedestrians in large part is due to uncertainty in AV intent and unfamiliarity with AV technology [147]. Once AVs' initial uncertainty

decreases, people may be more willing to interact with AVs and be more trusting and accepting of AVs. Thus, we hypothesized that once the initial uncertainty has been reduced after a few interactions with the AV, the increase in trust ($\Delta$ trust) between subsequent interactions will be higher in the case of AVs than for HDVs.

***H1**: Trust in human-driver will be **greater** than trust in AVs.*

***H2**: Change in trust between successive interactions will be **higher** for AV than human-driver.*

### 3.3.2 Method

For experimental consistency, we designed a situation and task similar to the one discussed in the VR user study (see Section 3.2.3.4). The experimental details are discussed below.

#### 3.3.2.1 Experimental setup and task

The study was conducted at the Mcity test track located in Ann Arbor. Mcity test track offers a safe, controlled environment to test AVs. The full-scale outdoor laboratory simulates the broad range of complexities vehicles encounter in urban and suburban environments and provides the connected infrastructure and operating system to serve as a smart city testbed. An urban portion of the track was used for this study. Crosswalks were marked in the middle of the street to simulate pedestrian crossings at unsignalized midblocks. Like the VR study, locations for picking up and dropping the ball were on opposite sides of a road. The path taken by the participants was similar to the VR study (refer Figure 3.7).

To maintain consistency, participants were asked to perform the same carrying task as in the previous VR study (refer Section 3.2.3. The difference is that instead of carrying a virtual ball, the participants carried numbered boxes from one side of the road to the other. Also, the participants had to move only two boxes. Thus, each participant crossed the street four times. In each of these crossings, the participant encountered an approaching vehicle (either an AV or an HDV) before crossing the street.

We manipulated the type of vehicle (automated or human-driven). However, both the vehicles were driven by the same person and followed a similar route and velocity profile (refer Section 3.3.2.4). We employed a between-subjects experimental design, so every participant experienced only one condition, i.e., they interacted either

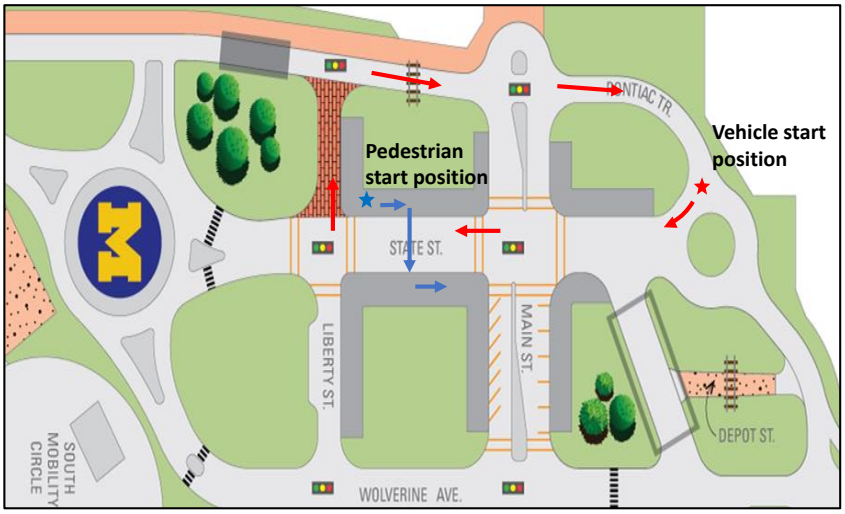



Figure 3.7: Urban portion of Mcity test track used for the study. The red star indicated the vehicle starting position, hidden from the pedestrian's sight initially. The blue star indicates the starting position of the pedestrian. The red and blue arrows indicate the paths followed by the vehicle and the pedestrian respectively for one interaction.

with the AV or the HDV. Participants were told there might be an AV running around the track but were not told the condition (human-driver or AV) they would be experiencing during the experiment. A sample scenario of pedestrian-AV interaction can be found online[2] for reference.

#### 3.3.2.2 Measurements

We used the same surveys as in the VR user study (refer Section 3.2.3.4) to measure participants' perceived driving aggression of AVs and trust in the AVs. Additionally, we asked the participants their perception of the vehicle's level of automation they encountered on a scale of 1-10, with 1 being 'no automation' and 10 being 'full automation'. Unlike the VR study, in this study, we have not measured the behavioral measures such as gaze, distance to collision, pedestrian speed, vehicle speed, jaywalking behaviors, etc.

[2]shorturl.at/xDEFV

#### 3.3.2.3 Study Participants

We recruited participants through email and obtained informed consent from each participant. Fourteen university students (4 female) joined in this study. We excluded three participants from the analysis as they failed the manipulation check (see Section 3.3.3). The remaining eleven participants had a mean age of 25.6 years (standard deviation = 3.5 yrs). This research complied with the American Psychological Association Code of Ethics and was approved by the institutional review board at the University of Michigan.

#### 3.3.2.4 Design of Interaction

Inspired by [63], we employed a WOZ methodology wherein the actual driver was disguised using a car seat costume as shown in Figure 3.8.

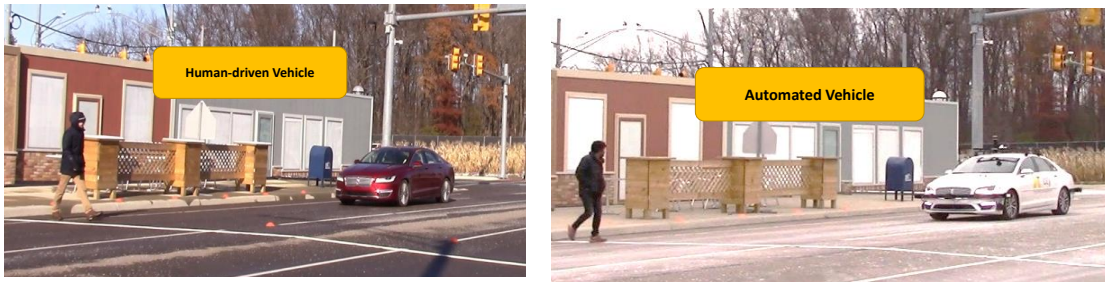

(a) (b)

Figure 3.8: WOZ driver inside the AV when **(a)** visible and **(b)** hidden.

The same driver drove both the AV and HDV and was trained to follow a similar velocity profile in both cases with a maximum speed of 15 mph. The vehicle was initially out of sight from the pedestrians indicated by the vehicle start position in Figure 3.7. The interactions between pedestrians and the vehicles were timed based on cues given by the experimenter to the driver using walkie-talkies.

### 3.3.3 Results and discussion

We performed a manipulation check to test if pedestrians perceived the WOZ AV to be an actual AV (based on the reported perceived level of automation). Seven out




(a) (b)

Figure 3.9: User study interaction conditions **(a)** a pedestrian crossing the street in front of an oncoming HDV, and **(b)** a pedestrian crossing the street in front of an oncoming AV.

of eight participants in the AV condition thought the vehicle to be an AV, and the remaining one participant could not see that there was no person inside the vehicle due to glare. Four out of six participants in the HDV condition saw the human and identified the vehicle as an HDV. The remaining participants did notice the human inside the vehicle but instead assumed the vehicle to be an AV since the study was conducted at the Mcity test track. Still, we removed the two participants' data who did not correctly identify the vehicle type from the analyses.

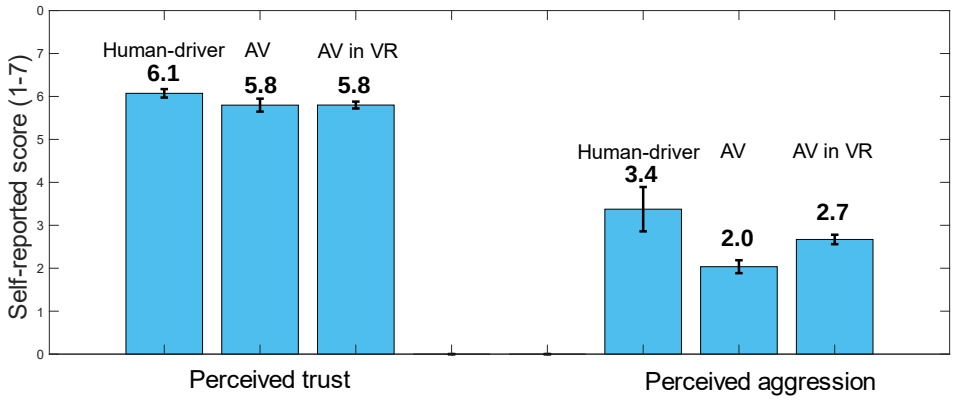



Figure 3.10: Pedestrians trusted the HDVs more than the AVs. However, pedestrians thought the HDVs to be more aggressive than the AVs.

The sample size (N=11, four from HDV condition and seven from AV condition) is small to obtain a statistical significance. Hence, we checked the trends in the data to test our hypotheses. To test H1, we compared the mean self-reported trust of the participants for both the HDV and AV conditions, and we found that pedestrians trusted the HDVs, $\overline{trust}_{hdv} = 6.1$, more than the AVs, $\overline{trust}_{av} = 5.8$) (see Figure 3.10). Thus H1 is supported.

We also found that the change in trust in the subsequent interactions with AVs ($\Delta$ trust = 0.35) was higher than that in the subsequent interactions with HDVs ($\Delta$ trust = 0.24). The results suggest that pedestrians could be skeptical of AVs and do not trust the AVs initially, which aligns with the findings from AV acceptance surveys [147]. However, as the uncertainty in AV behavior reduces with subsequent AV interactions, the trust in AV increases. Thus H2 is also supported.

To compare the pedestrian perception of AVs from the Mcity study with the VR study, we compared the trust in the AVs and the perceived aggression of both the studies (refer Figure 3.10). Trust in the AVs was essentially the same for both the VR and the real-world Mcity studies. However, the testing conditions in both studies were not the same. For example, the behaviors of the AVs in the VR and Mcity were different, and the VR had three different behaviors while the Mcity had only one behavior. This can also be observed by the different perceived aggression for the AV driving in the VR (aggression = 2.7) and the Mcity (aggression = 2.0) studies. Another factor is the number of participants, which was considerably higher in the VR study than in the Mcity study. Still, the similarity in the trust levels could indicate underlying similarities in the perception of the pedestrians towards AVs irrespective of the type of the testing environment. Further research is necessary for any conclusive findings.

Another interesting observation we found was that the pedestrians considered the HDVs to be more aggressive than the AVs (refer Figure 3.10). However, they still trusted the HDVs more than the AVs. This suggests that the pedestrians could be more forgiving of the HDVs, even if they are more aggressive than the AVs. It also suggests that the pedestrians are expecting the AVs to be less aggressive. Thus directly imitating human-like behavior by the AVs [149] may not be beneficial in improving traffic safety as pedestrians are expecting the AVs to be more conservative than the HDVs. These behavioral expectations from the AVs could inform the AV planner on how the AVs should behave around the pedestrians.

## 3.4 Chapter summary

In this chapter, we characterized the role of AV driving behavior on pedestrian trust in the AVs and pedestrian behaviors; the results have been published in [33, 34]. In addition to confirming existing research that AV intent can be deduced from their behaviors [18], we found that pedestrian trust and behavior are influenced by AV driving behaviors. However, this influence is valid only at unsignalized crosswalks. We

also identified relations between pedestrian's trust in the AVs and several behavioral (of both vehicle and pedestrian) and environmental factors. Interestingly, we also found that the novelty of AVs causes the pedestrians to have different expectations compared to the HDVs.

These findings aid in the development of pedestrian behavior models that can be used by the AVs for predicting pedestrian behavior discussed in Chapters IV and V.

# CHAPTER IV

# Individual Pedestrian Behavior Modeling

## 4.1 Overview of chapter

Current pedestrian models for long-term multimodal predictions are data and computation intensive and also lack intuition. The contribution presented in this chapter is the development of an intuitive and computationally efficient pedestrian behavior model that is suitable for long-term multimodal predictions; these results have been published in [35, 36]. We demonstrate the validity of the model on two datasets — one collected from the VR study discussed in Section 3.2.3.4 and the other is the inD dataset that captured pedestrian-HDV interactions in the real-world [37].

As noted in Chapter III, AV driving behavior can influence pedestrian trust in AVs. The AV can choose trustworthy trajectories if it can understand the intent of the pedestrians. This chapter develops a hybrid automata-based modeling framework that uses pedestrians' gap acceptance behavior and constant velocity dynamics for long-term unimodal pedestrian trajectory prediction when interacting with AVs. This model produces the most likely pedestrian trajectory. However, pedestrian behavior is uncertain in nature. To account for this uncertainty, we extend the proposed hybrid systems framework to predict multimodal pedestrian behavior.

We focus our efforts in modeling pedestrian behavior around unsignalized crosswalks present at midblocks and intersections. The right-of-way is unclear at such unsignalized crosswalks [32] which can result in potentially dangerous interactions between AVs and pedestrians. Models were developed and validated in datasets containing pedestrian-AV interactions in VR and pedestrian-HDV interactions in the real world. Results demonstrate the applicability of both the unimodal and multimodal models for long-term ($> 5\,s$) pedestrian trajectory prediction at unsignalized crosswalks.

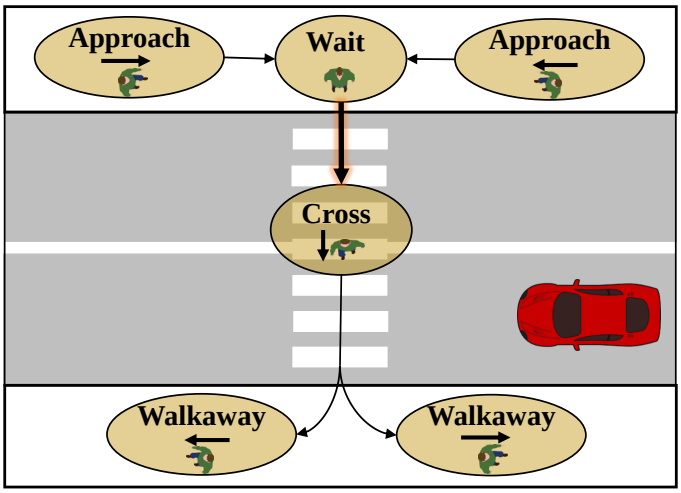



Figure 4.1: Pedestrian behaviors when they intend to cross the road. Pedestrians are assumed to use the crosswalk for crossing and can be doing one of four actions at any given time – approaching the crosswalk, waiting near the crosswalk (and deciding when to cross), crossing, or walking away from the crosswalk. The bubbles represent the actions, and the arrows represent the action transitions. The bold arrow represents the transition from wait to cross, i.e., the pedestrian's decision to cross.

## 4.2 Pedestrian behavior in urban environments

Pedestrians can engage in many behaviors near crosswalks. For example, some pedestrians may intend to cross while some may not intend to cross, and some others can even jaywalk instead of using the crosswalk to cross. The key idea we employ in this model is that pedestrian behaviors are driven by their actions; actions are driven by their decisions which in turn depends on their intents. By using pedestrians' intents, decisions, and actions, we should get a good sense of their behaviors.

Let us revisit our definition of pedestrian behavior discussed in Section 1.1, "*Pedestrian behavior* refers to the actions and, in turn, the paths (trajectories) taken by pedestrians." Generally around crosswalks, pedestrians can be doing one of three actions: (1) *approach crosswalk*, (2) *wait*, or (3) *cross* [150]. Additionally, pedestrians can also be doing a fourth action, (4) *walkaway* from the crosswalk, after crossing the road. Figure 4.1 shows these distinct actions and the more common behaviors (indicated by the arrows) of pedestrians intending to cross the street. Thus, a pedestrian can be considered a hybrid system with discrete actions and continuous motion within each action.

The trajectory traversed by the pedestrian is a result of the intermediate decisions made by the pedestrian. These decisions could be taken within a particular action,

such as walking speeds of pedestrians while approaching or crossing, or the decisions could be taken to transition from one action to another, such as accepting a traffic gap and changing the action from wait to cross. Here, we focus on the latter type of decisions, i.e., decisions that change the actions. A hybrid systems model is developed based on this decision-making framework which is discussed in the coming sections.

We assumed that pedestrians are either walking or standing still. If they are walking, they can be approaching the crosswalk, crossing, or walking away from the crosswalk. They always walk only on the sidewalk, or across the crosswalk, at a constant velocity. If they are standing still near the crosswalk, they are waiting to cross, and they always use a crosswalk to cross the street.

## 4.3 Unimodal hybrid pedestrian model

We propose to use a hybrid automata framework [151] to model the hybrid systems nature of pedestrian behavior. The hybrid automaton has four discrete states—***approach crosswalk, wait, cross, and walk away***—each with an associated continuous motion. The sequence of these discrete states and the state transitions, together with the pedestrian trajectory over time, gives rise to the pedestrian's crossing behavior. The transition from wait to cross state is modeled as pedestrian's decision to cross, which is explained in section 4.3.1. The proposed hybrid automaton model is shown in Figure 4.2 and is formally defined as a tuple $\langle X, Q, f, G, T, R \rangle$, where

- $X = (x, y, v_x, v_y) \in \mathbb{R}^4$, denotes the continuous pedestrian states which includes their Cartesian positions $(x, y)$ and velocities $(v_x, v_y)$.
- $Q \in \{q_1, q_2, q_3, q_4\}$, denotes the discrete states of approach, wait, cross, or walk-away respectively.
- $f : Q \times X \rightarrow \mathbb{R}^4$, represents the continuous dynamics.
- $G \in \{g_1, g_2, g_3, g_4\}$ is a set of guard conditions, where
  - $g_1 = \{X \mid v_x \neq 0 \wedge sign(x)\, v_x < 0\}$
  - $g_2 = \{X \mid v_x = 0 \wedge v_y = 0\}$
  - $g_3 = \{X \mid v_y \neq 0\}$
  - $g_4 = \{X \mid v_x \neq 0 \wedge sign(x)\, v_x > 0\}$.

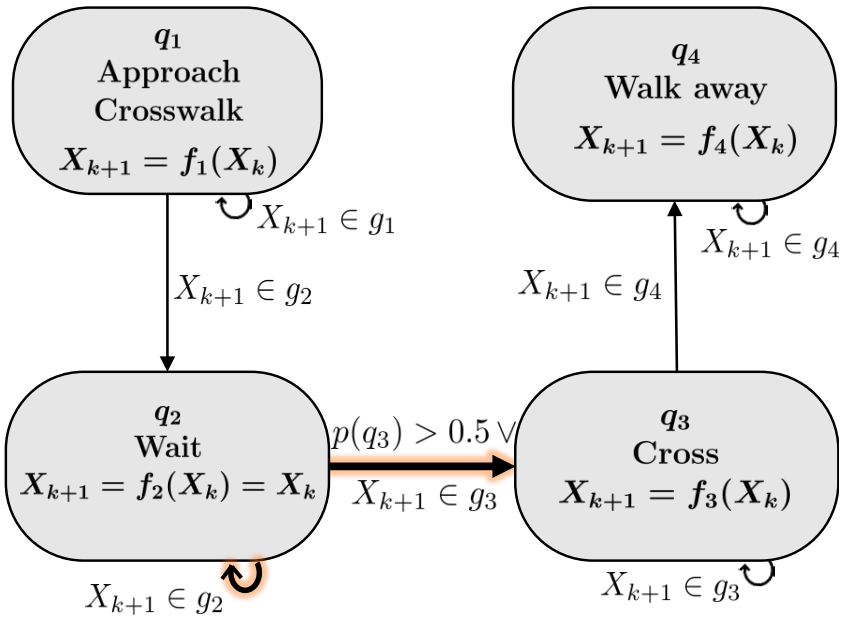



Figure 4.2: Hybrid automaton of a rational pedestrian with the intent to cross. The bold arrows represent the transitions from the crossing decision model.

- $T$ is the discrete state transition function given as,

  - $T(Q_{k+1}|Q_k) = \begin{cases} q_1 & \text{if } X_{k+1} \in g_1 \\ q_2 & \text{if } X_{k+1} \in g_2 \\ q_3 & \text{if } X_{k+1} \in g_3 \vee p(q_3) > 0.5 \\ q_4 & \text{if } X_{k+1} \in g_4 \end{cases}$

    where $p(q_3)$ is the probability of crossing.

- $R : X \to X$ is a reset map of continuous velocity states after the discrete transitions.

The developed model outputs a single trajectory that is predicted to be the most likely future trajectory. We term this model as **Unimodal hybrid pedestrian (UHP)**. We make the following assumptions for developing the UHP model.

**Assumption IV.1.** *Pedestrian behavior is primarily influenced by its interaction with the vehicle that is approaching and is closest to the pedestrian, referred to as the* interaction vehicle *(InterV).*

**Assumption IV.2.** *The position and shape of the crosswalks and sidewalks are accurately known.*

This model focused on the interaction between the pedestrian and the InterV. When the current InterV moves past the pedestrian, the interaction is then modeled with the next approaching vehicle, which then becomes the InterV.

**Assumption IV.3.** *All of the observations and predictions are made from the point of view of the vehicle termed as the* ego-vehicle *(EgoV).*

In this dissertation, we are considering the ego-vehicle to be an AV.

**Assumption IV.4.** *The EgoV's observations of the other vehicles, pedestrians, and itself can be measured with known and bounded sensor noises.*

**Assumption IV.5.** *All pedestrians approaching a crosswalk have the intent to cross.*

This is a conservative assumption but simplifies the problem of pedestrian behavior modeling.

**Assumption IV.6.** *Within a discrete state, a pedestrian moves at a constant velocity.*

As mentioned before, pedestrians can also make decisions within an action, resulting in distinctly different dynamics. We, however, neglect these decisions and assume that pedestrians travel at a constant velocity while performing an action. Prior work has found that constant velocity dynamics has worked well for long durations provided there are not many changes in the direction of the pedestrians [152, 153].

**Assumption IV.7.** *Pedestrians always use the crosswalk for crossing the road, i.e., pedestrians do not jaywalk.*

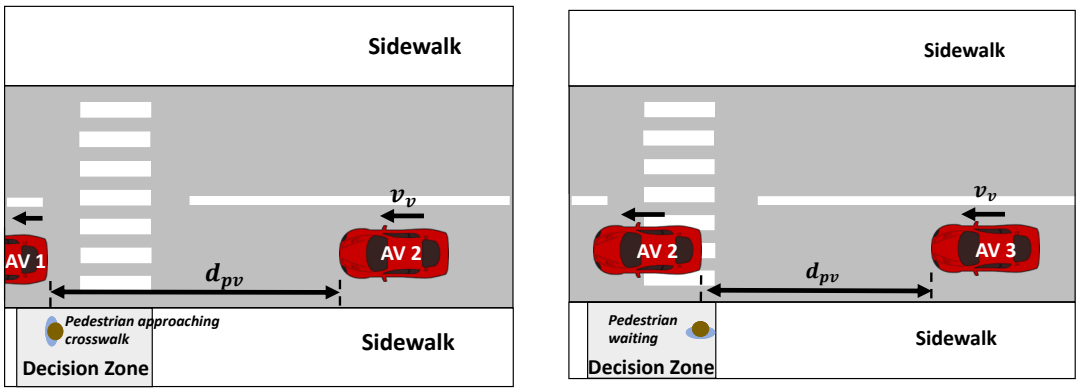
**Assumption IV.8.** *Gaps are accepted only when the pedestrian is close to the crosswalk and is either approaching the crosswalk or waiting.*

This region of gap acceptance is denoted by the decision zones, D, in Figure 4.3.

### 4.3.1 Crossing decision model

We describe pedestrian discrete state transition of *wait to cross* (i.e., crossing decision) through their gap acceptance behavior [43]. We develop a model that outputs the probability of accepting a traffic gap. Since gap acceptance is a discrete phenomenon, we make the following assumptions.

**Assumption IV.9.** *The decision to accept/reject a gap is made at the start of a gap, and the decision holds for the entire duration of that gap.*




(a) (b)

Figure 4.3: Evaluation of gap acceptance: **(a)** a pedestrian is approaching and close to the crosswalk and a gap starts, and **(b)** a pedestrian is waiting on the road and a gap starts. Pedestrians' decide to accept/reject the gaps when they are in a decision zone, D.

**Assumption IV.10.** *A gap starts when the pedestrian is close to the crosswalk and is in the decision zone D. A vehicle has just gone past the pedestrian, exposing the pedestrian to interact with the next oncoming vehicle (refer Figure 4.3).*

**Assumption IV.11.** *We assume that pedestrians always enter the wait state before crossing irrespective of whether they had stopped at the crosswalk before crossing or not.*

The wait is instantaneous when the pedestrian has decided to cross while in the approach state. The pedestrians enter the "wait" state for an infinitesimal time and immediately transition to the "cross" state. The discrete state transitions are triggered when the corresponding guard conditions ($G$) are satisfied. Additionally, the *wait to cross* transition can also be triggered by the decision to cross ($p(q_3) > 0.5$).

**Assumption IV.12.** *Gaps are evaluated when pedestrians are in the "wait" or "approach crosswalk" state and within the decision zone.*

We model crossing decision using a Support Vector Machine (SVM) classifier and obtain probabilistic outputs following the method in [154]. The input parameters to the SVM model were identified from literature [155, 156] as significantly affecting pedestrian crossing behavior. They are detailed in Table 4.1 and are denoted by $\psi_i$, where $i$ is the time step.

Table 4.1: Input parameters ($\psi$) for crossing decision model. The parameters are calculated for the current time instance when a gap starts unless mentioned otherwise.

| **Parameter** | **Description** |
|---|---|
| AV distance $[m]$ | Longitudinal distance between AV and pedestrian |
| AV speed $[m/s]$ | Speed of the AV |
| Wait time $[s]$ | Time elapsed since pedestrian started waiting |
| Gaze ratio | Proportion of time pedestrian looked at AVs in previous second |
| Curb distance $[m]$ | Lateral distance between pedestrian & road edge |
| Crosswalk distance $[m]$ | Longitudinal distance between pedestrian & crosswalk |
| Ped. speed $[m/s]$ | Average pedestrian speed in the previous second |

### 4.3.2 UHP model for real-time trajectory prediction

We express the pedestrian continuous dynamics using a constant velocity model with zero-mean Gaussian process noise ($W$) as $X_{k+1} = f(X_k, \mathcal{N}(0, W))$. We incorporate the constant velocity continuous dynamics within a Kalman filter framework and the SVM crossing decision model in our hybrid automaton model for tracking pedestrians' position in real time. We tune the process and measurement noises to obtain the best tracking performance. We used MATLAB's in-built *tune* function from the Navigation toolbox developed based on [157] to tune the noise parameters. The UHP model parameters are tabulated in Table 4.2.

Table 4.2: UHP model parameters.

| **Parameter** | **Value** |
|---|---|
| Decision zone length, $D$ | $3\,m$ |
| Prediction horizon, $N$ | $0.1 - 6\,s$ |
| Sampling time, $\Delta t$ | $0.1\,s$ |
| sensing range of the AV | $100\,m$ |

We evaluate the UHP model by predicting pedestrian trajectories. Algorithm 1 shows the steps for real-time pedestrian trajectory prediction. When the pedestrian is within the sensing range of the AV, the vehicle and pedestrian measurements are used to calculate the initial pedestrian states and the initial value of the features (refer Table 4.1). The inference framework has two stages, predict and update. The pedestrian motion is predicted using the continuous motion model during the predict stage, and the discrete state is updated based on which guard condition the continuous state satisfies. Additionally, the transition from wait to cross can be predicted using the SVM crossing decision model. The decision to cross is evaluated when the pedestrian is within the decision zone and when a gap starts. The prediction loop continues for every time step within the prediction horizon $N$, using the previously

predicted measurements. During the update stage, the states are updated based on the prediction for that time instant and the measurement.

**Algorithm 1** Inference framework for Hybrid model

Initialize $X_1, Q_1, \psi_1, i \leftarrow 1$
**while** $t_i < T$ **do**
  **procedure** PREDICT
    ▷ Predict discrete and continuous states for entire prediction horizon
    **for** $k \leftarrow 1$ to $N$ **do**
      **if** (Gap Starts $\wedge\, (\hat{X}_{i+k} \in D) \wedge \hat{Q}_{i+k} == q_2$) **then** ▷ check gap acceptance probability
        Calculate $P(q_3) = P(\hat{Q}_{i+k+1} = q_3 | \hat{\psi}_{i+k+1})$
      **end if**
      Sample $t_{cross}$, $v_{start}$ if $P(q_3) > 0.5$
      $\hat{Q}_{i+k+1} \leftarrow h(\hat{X}_{i+k}, P(q_3), t_{cross})$
      Reset $\hat{X}_{i+k}$ if $\hat{Q}_{i+k+1} \neq \hat{Q}_{i+k}$
      $\hat{X}_{i+k+1} \leftarrow f(\hat{X}_{i+k}, \hat{Q}_{i+k+1})$
      Estimate $\hat{\psi}_{i+k+1}$
      $k \leftarrow k + 1$
    **end for**
  **end procedure**
  **procedure** UPDATE
    $X_{i+1} \leftarrow$ update state given $\hat{X}_{i+1}$, measurement $z_{i+1}$
    $Q_{i+1} \leftarrow h(X_{i+1})$
  **end procedure**
  $i \leftarrow i + 1$
**end while**

There is a delay between deciding to cross and the start of crossing [158]. We express this delay as $t_{cross}$ and sample it from an exponential distribution learned from the collected data. The transition from wait to cross occurs when it is $t_{cross}$ seconds since the time gap was accepted (i.e., a time delay is reached). Similarly, the pedestrian speed when starting to cross, $v_{start}$, is sampled from a Gaussian distribution learned from the data. During the prediction stage, we assume constant velocity dynamics for the AVs. When no measurement is available, the features $\psi_i$ within the prediction horizon are calculated from the predicted pedestrian and AV positions. We assume that the gaze ratio of pedestrians remains the same for the entire prediction horizon as the most recent observation. Using predicted data instead of actual measurements for the features, the hybrid model can perform long-term trajectory prediction.

### 4.3.3 Baseline models

We compared the crossing decision model with two baselines—a logistic regression model, similar to [43], and a model trained using the conditional probability

distributions of gap acceptance conditioned on the observations mentioned in Table 4.1.

Current collision avoidance systems use a constant velocity (CV) model for pedestrian trajectory prediction [159]. Thus, similar to [21], we compared our hybrid model against a baseline constant velocity model without any discrete states. We report the trajectory prediction performance at varying prediction horizons.

### 4.3.4 Evaluation metrics

We use the following metrics to compare the performance of the crossing decision model with baseline classification models.

***Accuracy***: ratio of correctly predicted observation to the total samples.

***Precision***: ratio of correctly predicted positive ('cross decision') samples to the total predicted positive samples.

***Recall***: ratio of correctly predicted positive ('cross decision') samples to the actual overall positive samples.

***F1-Score***: this metric balances both precision and recall.

The metrics are calculated as,

$$Accuracy = \frac{TP + TN}{NS}, \tag{4.1}$$

$$Precision = \frac{TP}{TP + FP}, \tag{4.2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{4.3}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4.4}$$

where TP is the true positives, FP is the false positives, FN is the false negatives, TN is the true negatives, and NS is the total number of samples.

We use the following metrics to compare the performance of trajectory prediction from the UHP model with a baseline trajectory prediction model.

***Average displacement error (ADE)***: mean distance between predicted and actual trajectories for all time steps in the prediction horizon.

***Final displacement error (FDE)***: distance between predicted and actual position at the end of prediction horizon.

Consider the ground truth position at time step $i$ to be given by $(x_i, y_i)$. Similarly, the predicted position corresponding is given by $(\hat{x}_i, \hat{y}_i)$. The ADE and FDE are thereby given as,

$$ADE = \frac{\sum_{i=1}^{N} \|(x_i, y_i) - (\hat{x}^i, \hat{y}^i)\|^2}{N}, \tag{4.5}$$

$$FDE = \|(x_N, y_N) - (\hat{x}_N, \hat{y}_N)\|^2. \tag{4.6}$$

### 4.3.5 Dataset description

We used the dataset collected from the VR user study (refer Section 3.2.2), termed as the automated vehicle interaction in virtual reality (AVIVR) dataset, for the development of the crossing decision model and the evaluation of the trajectory prediction performance of the UHP and baseline models. The dataset was collected at an unsignalized midblock crosswalk with one-way traffic of AVs. The dataset contained trajectory information of the AVs and pedestrians and the gaze information of the pedestrians from the eye tracker. The trajectory information of the AVs is obtained directly from the simulation (i.e., perfect information, which is unlikely in the real world). The pedestrian trajectories are also recorded from the simulation, but the inputs to the simulation are through the IMU sensors in the shoes and the ring surrounding the pedestrian (i.e., study participant, refer Section 3.2.2.1). Thus pedestrian trajectory data would include noise from the IMU sensors. Gaze data from the eye tracker provides information on the region(s) the pedestrian is looking at every sampling instant. This is done by obtaining the gaze point from the eye tracker and casting a ray to identify which region intersected with the gaze ray automatically (refer Section 3.2.2.5). Gaze point identification can be noisy and only data with a minimum confidence level (provided by the Pupil Labs software) of 0.90 was used for gaze region identification.

In this dataset, there was always exactly one pedestrian at any given time. Each pedestrian always crossed the street. So thirty participants for three treatment conditions with six crossings at unsignalized scenarios resulted in 540 crossing pedestrians trajectories. The dataset was processed to identify traffic gaps, both accepted and

rejected gaps, for each of the 540 crossings. Instances of traffic gaps were identified when a pedestrian was inside the decision zone and when a traffic gap has just started. The model features (refer Table 4.1 were compiled for these time instances. After cleaning the gap data and removing outliers, the dataset contained 508 crossing trajectories with 508 accepted gaps and 1195 rejected gaps. The gap data were used for developing the crossing decision model, and the trajectory data was used for evaluating the UHP model. The dataset was split into training (80 %) and testing (20 %) sets, i.e., 406 crossing trajectories and accepted gaps and 934 rejected gaps for training and 102 crossing trajectories and accepted gaps, and 261 rejected gaps for testing.

### 4.3.6 Results and discussion

#### 4.3.6.1 Crossing decision

We used the SVM gap acceptance model for predicting the crossing decision. The training gap data is used for model development. Since the dataset is imbalanced, i.e., more rejected gaps than accepted gaps, we performed oversampling and resampled gaps from the available accepted gaps. After testing the performance of various kernel functions (ranging from the simple kernels to more complicated kernels) for the SVM classifier—linear, quadratic, cubic, Gaussian—we chose the cubic Kernel for the SVM model since it had the highest performance in terms of F1-score.

We compared the model with two baselines—a logistic regression model. We used F1-score for model comparison. As shown in Table 4.3, the SVM model performs better than the baselines and is used for pedestrian trajectory prediction. Table 4.4 shows the importance of the various parameters based on the SVM model performance [160]. The features are arranged in descending order of importance. The gaze ratio has the least impact on the model performance.

Table 4.3: Comparison of crossing decision models.

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Probability Distributions | 0.76 | 0.66 | 0.51 | 0.56 |
| Logistic Regression | 0.79 | 0.71 | 0.52 | 0.60 |
| Support Vector Machine | 0.88 | 0.75 | 0.73 | 0.74 |

Table 4.4: SVM crossing decision model feature ranking.

| Feature removed | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| AV distance | 0.85 | 0.74 | 0.55 | 0.63 |
| Crosswalk distance | 0.86 | 0.71 | 0.66 | 0.69 |
| Curb distance | 0.86 | 0.71 | 0.71 | 0.71 |
| Wait time | 0.87 | 0.72 | 0.72 | 0.72 |
| AV speed | 0.88 | 0.78 | 0.68 | 0.73 |
| Ped. speed | 0.87 | 0.74 | 0.71 | 0.73 |
| Gaze ratio | 0.88 | 0.76 | 0.75 | 0.75 |

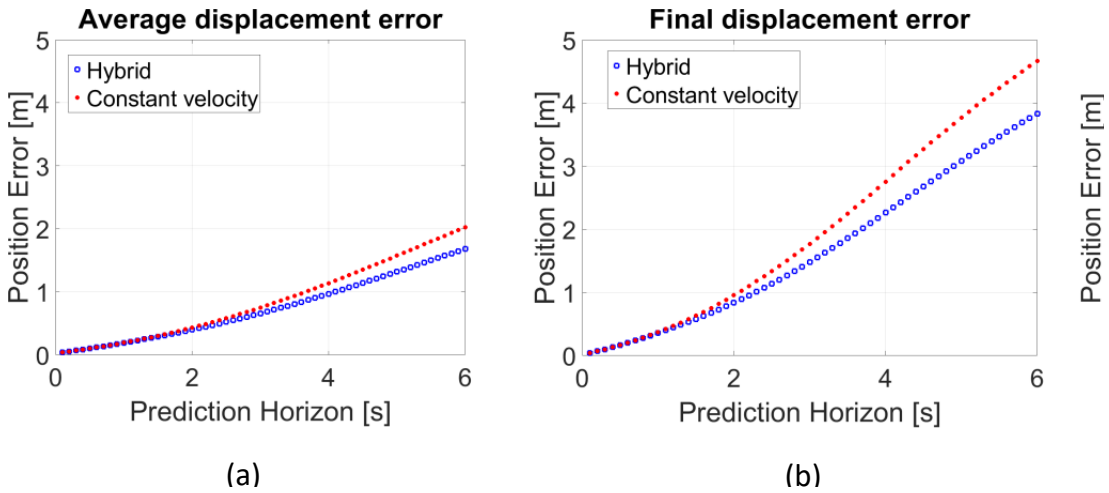



Figure 4.4: Pedestrian tracking comparison for different prediction horizons for (a) Average displacement error, and (b) Final displacement error. The hybrid model has lower error than the constant velocity baseline across both the metrics.

#### 4.3.6.2 Hybrid Model Real-time Trajectory Prediction

The developed UHP model was used for predicting pedestrian crossing trajectories. The transition from wait to cross was predicted using the SVM crossing decision model. From Figure 4.4, it can be seen that the hybrid model performs better than the constant velocity model across all metrics. At short prediction horizons ($< 2\,s$), the models are similar in performance, with the performance difference between the UHP and CV models increasing with the prediction horizon. This demonstrates the applicability of the UHP model at long prediction horizons. The longer horizons typically capture the intents and the decisions.

We aimed at developing a model to characterize long-term pedestrian crossing behavior, which in turn can be used for long-term pedestrian trajectory prediction. The hybrid systems UHP model we developed incorporates pedestrian crossing decision-making and accounts for pedestrians already waiting at the crosswalk. This makes the

model suitable for pedestrian behavior (trajectory) prediction at crosswalks. Trajectories predicted using our hybrid model had lower errors than the baseline model, and the performance difference increased with the prediction horizon. Thus our model is also better suited for longer prediction horizons than existing models [21, 22]. Further, our UHP model includes the contextual information of the vehicle behavior (through vehicle distance and speed).

The SVM crossing decision model we developed can function with only the pedestrian's position information. In contrast, many existing models [22, 46] additionally require rich pedestrian pose information as discussed in Section 2.3. We should acknowledge that unlike previous studies [155, 161], we did not find a substantial relationship between pedestrian crossing and gaze behaviors. Including gaze ratio did not substantially improve the crossing decision predictions (refer Table 4.4). This could be because any rational pedestrian intending to cross can be expected to always look for the vehicles irrespective of their decision to cross or not.

We modeled pedestrian behavior around an unsignalized midblock crosswalk. Pedestrian behaviors approaching the crosswalk and at the crosswalks were primarily captured and modeled. The modeling framework should be generalizable to similar scenarios with an unsignalized midblock crosswalk as their approach, and waiting behaviors would be similar.

A limitation of the UHP model is that we assumed that all pedestrians had the intent to cross the street, which was valid for the user study we conducted but not applicable in the real world as pedestrians can be walking on the sidewalk without ever intending to cross. Further, it gives a single deterministic output trajectory, whereas pedestrian behaviors can be multimodal. We address these concerns in Section 4.5 by extending the model to predict multimodal pedestrian behavior.

Another limitation is that the model was developed on a dataset obtained from a controlled simulated environment, limiting its applicability to real-world conditions. The framework, however, can be extended to the real world. Pedestrian data is normally captured using perception systems such as cameras and Lidar in the real world. Measurement noises from these sensors can be incorporated in the Kalman filter of the UHP model in place of the IMU sensor noise currently used.

## 4.4 Pedestrian Behavior Comparison

The AVIVR is a good first step to collect a pedestrian-AV interaction dataset. However, the dataset is collected in VR, which may not bring out actual pedestrian

behavior as it does not have real-world risks and goals. On the other hand, there are several publicly available datasets with rich pedestrian-HDV interaction data [37, 40, 162, 163]. Current research utilize these datasets for developing pedestrian models for AV planning applications [28, 164]. However, they assume that pedestrian models derived from pedestrian-HDV interaction data are applicable for AV planning, which is not validated, given that pedestrians are likely to behave differently towards AVs (refer Section 3.1).

Collecting pedestrian-AV interaction data in the real world is also not feasible currently as it raises safety concerns. We compared some pedestrian behavioral measures estimated from the AVIVR dataset and published results from the real-world interactions. We compared two measures, gap acceptance and walking speed, describing crossing behavior during the wait to cross transition and during approach and cross states, respectively, with published results from real-world studies [43, 165]. For comparison validity, we chose real-world studies that had a similar road structure (two-lane uncontrolled mid-block crossing) as our IVE.

### 4.4.1 Gap Acceptance

We compared gap acceptance behavior in AV interactions with [43] and used the same gap measure as [43]. *Accepted traffic gaps* were the difference between two time points: the time when the pedestrian just stepped onto the road and the time when the head of the vehicle had just passed the pedestrian's longitudinal position. The comparison of the cumulative gap acceptance distributions is shown in Fig. 4.5 (a). We calculated KL-divergence [166] to compare the curves. KL divergence is a commonly used metric in ML and statistics to identify the divergence between two distributions based on information entropy. KL divergence was calculated for the cumulative distribution of gaps. KL divergence is normally used to identify the closest distribution among a set of distributions to the actual distribution. If the KL divergence is 0, it means the distributions provide the same information. Thus for the two distributions to be close, we expect the KL divergence to be small. We found the KL-divergence value to be 0.17, which is low [166].

### 4.4.2 Walking Speed

We calculated pedestrian speed as the finite difference of their positions and applied a moving average filter to reduce noise. Pedestrians tend to walk faster while crossing ($1.58\,m/s$) [165] than on sidewalks ($1.48\,m/s$). We observe a similar trend in

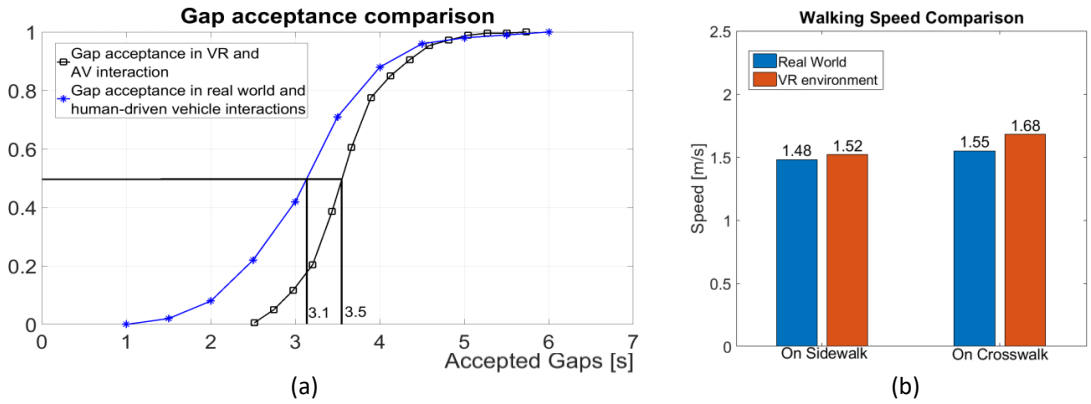



Figure 4.5: Comparison of measures of crossing behavior. (a) Similar cumulative probability curves for gap acceptance in both AV and HDV scenarios [43]. (b) Higher walking speeds observed while crossing than on sidewalk, in both AV and HDV cases [165].

pedestrian speeds in AV interactions ($1.68\,m/s$ while crossing compared to $1.52\,m/s$ while on sidewalk), as shown in Fig. 4.5 (b).

### 4.4.3 Discussion

We examined and found some similarities between measures of crossing behavior, namely, gap acceptance and walking speed, during interactions with AVs in VR and HDVs in the real world. These results are in line with [63, 68], where pedestrians behaved normally around AVs as they would around HDVs. The similarities suggest two possibilities: (i) the hybrid models developed from the VR dataset can be applicable for real-world scenarios with AVs, and (ii) the publicly available pedestrian-HDV datasets can be applicable for AV planning applications. However, further investigation is needed to validate these possibilities.

## 4.5 Multimodal hybrid pedestrian (MHP) Model

Pedestrian behavior is multimodal in that pedestrians can take multiple possible actions that result in multiple different trajectories in any given situation. For example, consider a pedestrian ($P_1$) approaching an intersection as shown in Figure 4.6. The pedestrian can either decide to cross at the nearby crosswalk or wait for a suitable gap in the oncoming vehicles to cross. Alternatively, they could cross at an adjacent crosswalk or walk away from the intersection altogether.

An AV can never be entirely sure what a pedestrian will do next; therefore, trajectory prediction is inherently stochastic. On the one hand, it is essential to consider

multiple possible pedestrian behaviors. On the other hand, not all possible pedestrian behaviors are relevant to AVs. For example, AVs should be more concerned about pedestrians who intend to cross and are about to cross than those just walking on the sidewalk. The trajectory prediction performance in such relevant situations must be satisfactory.

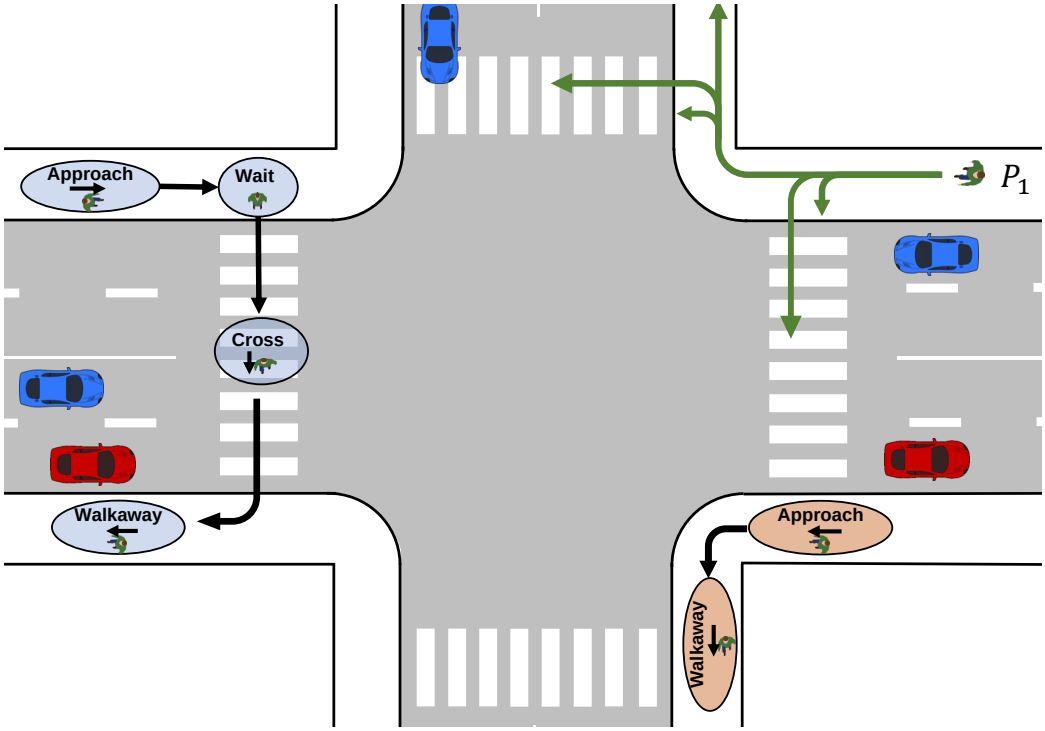



Figure 4.6: The figure shows some possible pedestrian behaviors at an intersection. Pedestrians, if they want to cross, are assumed to use the crosswalk for crossing. At any given time, pedestrians can be doing one of four actions – approaching the crosswalk, waiting near the crosswalk (and deciding when to cross), crossing, or walking away from the crosswalk. The bubbles represent the actions, and the arrows represent the action transitions. Pedestrian $P_1$ can potentially take multiple paths while approaching a crosswalk, as indicated by the green arrows. The blue bubbles indicate typical pedestrian behaviors when they intend to cross the road, and orange bubbles show typical pedestrian behaviors when they do not intend to cross the street.

We also relax the assumption that all pedestrians intend to cross the street. Instead, we separately modeled pedestrians' intent to cross and incorporated it into the trajectory prediction algorithm. Further, the hybrid systems framework is extended to unsignalized intersections in addition to midblocks. We define pedestrians as having the *intent to cross a crosswalk* if they either wait to cross or cross at that particular crosswalk. Figure 4.6 shows some possible differences in behavior for pedestrians with the intent to cross and for those not intending to cross. Pedestrians intending to

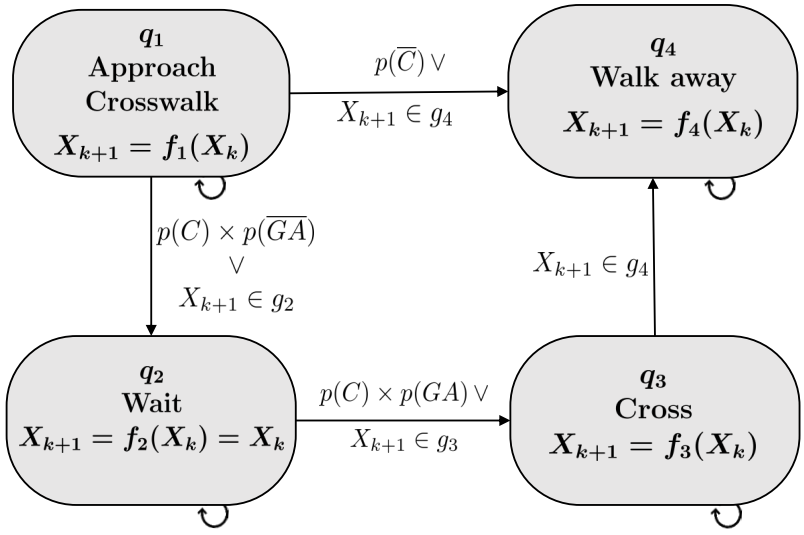


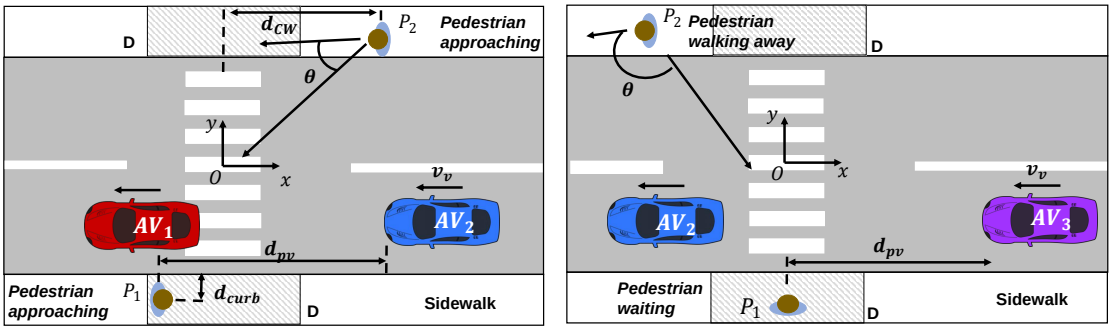

Figure 4.7: MHP, a probabilistic hybrid automaton model of a pedestrian. The transitions between states $q_1$ and $q_2$ and between $q_2$ and $q_3$ are determined by the predictions from the crossing intent model and crossing decision model.

cross normally evaluate the vehicle gap available and may wait by the sidewalk before starting to cross, whereas pedestrians not intending to cross may simply walk past the crosswalk or move to another crosswalk at the intersection. Pedestrian behaviors are not entirely independent but are influenced by the behavior of the vehicle approaching the crosswalk [16, 123]. We define this vehicle as the *interaction vehicle* (InterV), which is the vehicle closest to the pedestrian and approaching the pedestrian.

We extended the hybrid automata framework to develop the *Multimodal Hybrid Pedestrian* (MHP) model that can output multimodal predictions of pedestrian behavior. The MHP model utilizes individual crossing intent and crossing decision models in a probabilistic hybrid automaton framework for real-time prediction of pedestrian trajectories. We model the behavior of a single pedestrian interacting with one vehicle approaching the pedestrian.

Given a history of observations for $M$ time steps, denoted by $\hat{O}_{t-M:t}$, we would like to predict the pedestrians' trajectories $N$ time steps into the future given by, $\tilde{X}_{t+1:t+N} \in \mathbb{R}^2$. The observations for the $M$ time steps $\hat{O}_{t-M:t}$ include the trajectory of the pedestrians $\hat{X}_{t-M:t} \in \mathbb{R}^2$, the gaze of the pedestrians $\hat{G}_{t-M:t} \in \mathbb{R}^3$, and the trajectory of the interacting vehicle $\hat{X}^i_{t-M:t} \in \mathbb{R}^2$. Other information such as pedestrian and vehicle velocities and relevant distances are calculated from the observed pedestrian and vehicle trajectories (refer Tables 4.5 and 4.6). Pedestrians approaching a crosswalk at an intersection do not necessarily intend to cross it. They might

walk past the crosswalk and choose to cross at another crosswalk or walk away from the intersection. Similarly, once they are at the crosswalk and intend to cross, they can choose to cross immediately or wait for a safe vehicle gap. These intents and decisions of the pedestrians are inherently probabilistic and can result in multiple possible pedestrian behaviors and, in turn, multiple possible trajectories. For safe navigation, AVs must predict the more probable pedestrian trajectories but still be aware of all the possible pedestrian trajectories.




(a) (b)

Figure 4.8: **(a)** A pedestrian $P_1$ is approaching and close to the crosswalk and a gap w.r.t to $AV_2$ starts while another pedestrian $P_2$ is also approaching the crosswalk on the other side of the road, and **(b)** a pedestrian is waiting on the road and a gap starts while the other pedestrian walks away from the crosswalk without crossing. Pedestrians decide to accept/reject the gaps when they are in a decision zone, $D$.

Similar to the UHP model, the MHP model has the four distinct pedestrian actions as its discrete states—***approach crosswalk, wait, cross***, and ***walk away***—and each discrete state has an associated continuous motion. The different sequences of these discrete states and the associated pedestrian trajectory give rise to the different pedestrian behaviors. From the perspective of the ego-vehicle, some of these behaviors are more important than others. For example, when pedestrians are approaching the crosswalk, it is more critical that the AVs know when the pedestrians would cross and their crossing trajectory than their trajectory when they are on the sidewalk and approaching the crosswalk. The hybrid automaton framework we propose is generalizable to both midblocks and intersections and can model all possible pedestrian action sequences. However, unlike UHP, in this model, we do not assume all pedestrians have the intent to cross (assumption IV.5 is relaxed). Instead we model and predict the crossing intent of pedestrians. The MHP model focused on rational pedestrians who either cross at a crosswalk without jaywalking or do not intend to cross.

According to Assumption IV.6, similar to Keller and Gavrila [23], we express the discrete-time pedestrian dynamics within each discrete state as a constant velocity model with Cartesian position and velocity as the states, $X = (x, y, v_x, v_y)$ and zero-mean Gaussian process noise given as

$$X_{k+1} = f_i(X_k, \mathcal{N}(0, W)), \tag{4.7}$$

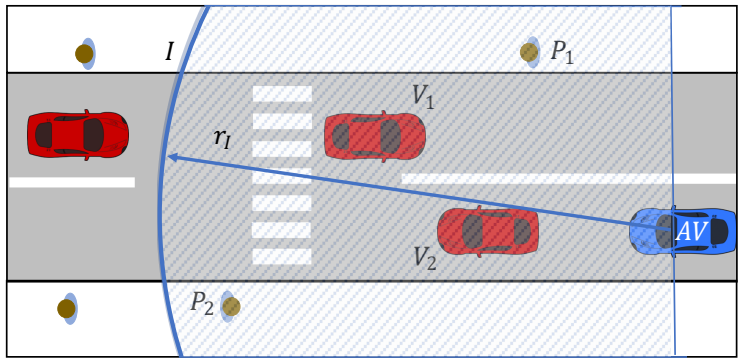
where $f_i$ represents the constant velocity function corresponding to discrete state $q_i$ and $W$ is the process noise. We model pedestrian dynamics using a point mass model.

The transitions between the discrete states are non-deterministic and probabilistic and depend on (i) the intent of the pedestrian to cross the street, (ii) the gap acceptance behavior of the pedestrians, and (iii) the guard conditions on the continuous states of the pedestrian that differentiate the discrete states. The pedestrians' crossing intent is predicted using a crossing intent model and their gap acceptance behavior is predicted using a crossing decision discussed in Sections 4.5.1 and 4.5.2, respectively.

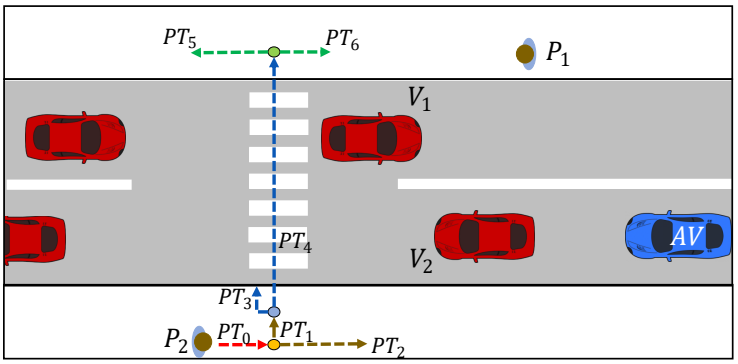
The proposed MHP model (shown in Figure 4.2) is formally defined as a tuple $\langle X, Q, f, G, T, R \rangle$, where

- $X = (x, y, v_x, v_y) \in \mathbb{R}^4$, are the continuous pedestrian states of Cartesian positions $(x, y)$ and velocities $(v_x, v_y)$.
- $Q \in \{q_1, q_2, q_3, q_4\}$, are the discrete states of approach, wait, cross, or walkaway respectively.
- $f : Q \times X \to \mathbb{R}^4$, represents the continuous dynamics.
- $G \in \{g_1, g_2, g_3, g_4\}$ is a set of guard conditions, where
  - $g_1 = \{X \mid v_x \neq 0 \wedge \theta \leq \theta_h\}$
  - $g_2 = \{X \mid v_x = 0 \wedge v_y = 0\}$
  - $g_3 = \{X \mid v_y \neq 0\}$
  - $g_4 = \{X \mid v_x \neq 0 \wedge \theta > \theta_h\}$,

  where $\theta$ is the angle difference between pedestrian heading angle and the angle between pedestrian and the crosswalk of interest (refer Figure 4.8) and $\theta_h$ is a threshold value for heading towards the crosswalk of interest.
- $T$ is the discrete state transition function given as,




(a)




(b)

Figure 4.9: **(a)** Within the interaction region $I$, AV predicts behaviors of pedestrians $P_1$, $P_2$ to their corresponding interactions vehicles, $V_1$, $V_2$ respectively **(b)** The different decision points along the trajectory of pedestrian $P_2$ are shown. The yellow, blue, and green nodes represent the decision points for crossing intent, crossing decision, and turn direction respectively and the colored arrows represent the different possible paths (prediction tracklets) the pedestrian can take at the decision points.

- $T(Q_{k+1}|Q_k) = \begin{cases} q_1 & \text{if } X_{k+1} \in g_1 \\ q_2 & \text{if } X_{k+1} \in g_2 \vee p(C) \times p(\overline{GA}) \\ q_3 & \text{if } X_{k+1} \in g_3 \vee p(C) \times p(GA) \\ q_4 & \text{if } X_{k+1} \in g_4 \vee p(\overline{C}), \end{cases}$

  where $p(C)$ is the probability of the pedestrian having the intent to cross and $p(GA)$ is the probability of the pedestrian accepting a gap and crossing the street.

- $R : X \to X$ is a reset map of continuous velocity and heading states after the discrete state transitions.

  The primary differences between the MHP and the UHP models are (i) the dis-

crete state transition function $T$, (ii) the guard conditions $G$, and (iii) the reset map $R$. In the case of the MHP model, the transition function depends on both the probability of the pedestrian having the intent to cross and the probability of accepting a gap instead of only the probability of accepting a gap in the UHP model. Unlike the UHP model, which was designed for a single unsignalized crosswalk, the MHP model was also applicable to an intersection scenario with multiple crosswalks. So, identifying the crosswalk the pedestrian is approaching is critical. Thus the guard conditions in the MHP model additionally depend on the heading of the pedestrians to identify if they are approaching the crosswalk of interest. The reset states, in general, depend on the crosswalk location and road orientation. Thereby, the reset states in the UHP were straightforward as there was only one crosswalk, while in the MHP, it varied based on the crosswalk of interest.

### 4.5.1 Crossing intent model

Pedestrians walking on the street may not always intend to cross the street. For example, in Figure 4.6, the pedestrian behavior depicted by the orange bubbles indicates that they do not intend to cross at the intersection. Current works mostly assume that all pedestrians have the intent to cross [22, 35, 106]. This, however, can result in highly conservative predictions of pedestrian behavior, which in turn can result in highly conservative AV behavior and cause the well-known freezing robot problem where the AV is unable to identify a safe trajectory to move forward.

We relax assumption IV.5 that all pedestrians have the intent to cross and instead model pedestrian's crossing intent. The ability to predict the crossing intent of pedestrians could help the AVs identify the most relevant pedestrians and aid in more effectively planning a path that is safe but not too conservative. We developed a model to predict the probability of pedestrians' intention to cross a street while approaching the crosswalk. We modeled this as a probabilistic classification problem using a support vector machine (SVM) [154].

Pedestrians intend to cross a street at a particular crosswalk based on their goal location. When they intend to cross at this crosswalk, they are likely to walk closer to the curb and slow down as they are approaching the crosswalk. This behavior would be independent of other factors such as the behaviors of the vehicles and the other pedestrians. Thus we use pedestrian-related features to model their crossing intent. We observed pedestrian trajectory, and gaze for an observation window of $3\,s$ (we tested observation durations of $1\,s$, $2\,s$, $3\,s$, and $4\,s$ and $3\,s$ gave the best perfor-

mance). The input parameters for the model were identified from literature [25, 50, 70] and are detailed in Table 4.5, denoted by $\lambda^m$ in Algorithm 2. Identifying crossing intent influences the discrete state transitions of *approach to wait* and *approach to walkaway*.

Table 4.5: Input parameters for crossing intent model ($\lambda^m$) calculated for the observation window of $3\,s$.

| **Parameter** | **Description** |
|---|---|
| Gaze ratio, $\phi_g$ | Proportion of time pedestrian gazed at InterV |
| Mean curb dist., $\bar{d}_{curb}$ $[m]$ | Average lateral distance between pedestrian & road edge |
| Mean crosswalk dist. $\bar{d}_{CW}$ $[m]$ | Average longitudinal distance between pedestrian & crosswalk |
| Mean ped. vel., $\bar{v}_p$ $[m/s]$ | Average pedestrian speed |
| Std. curb dist., $\tilde{d}_{curb}$ $[m]$ | Standard deviation of curb distance |
| Std. crosswalk dist. $\tilde{d}_{CW}$ $[m]$ | Standard deviation of longitudinal distance between pedestrian & crosswalk |
| Std. ped. vel., $\tilde{v}_p$ $[m/s]$ | Standard deviation of pedestrian speed |

### 4.5.2 Crossing decision model

To safely interact with crossing pedestrians, the AVs should predict whether and when pedestrians will cross the street. Pedestrians, when intending to cross the street, typically decide to cross the street by evaluating the safety of the available traffic gap from the interaction vehicle [43, 45]. A gap is considered to be accepted when the pedestrian starts crossing during that gap. The gap acceptance behavior influences the discrete state transitions of *approach to wait* and *approach/wait to cross*.

Similar to the crossing intent model, we modeled pedestrian's decision to cross as gap acceptance using a probabilistic support vector machine (SVM) classifier [154] that outputs the probability of accepting a traffic gap. Parameters that predominantly influence pedestrians' crossing decision were identified from the literature [16, 156] and used as inputs to the crossing decision model (refer Table 4.6), denoted by $\psi^m$ in Algorithm 2. Some of the parameters are illustrated in Figure 4.8.

### 4.5.3 MHP model implementation for real-time trajectory prediction

For predicting pedestrian trajectories in real time, we incorporate the individual models—crossing intent model, crossing decision model, and constant velocity continuous dynamics model (equation 4.7)—into the hybrid automaton framework described earlier (refer Figure 4.7).

Table 4.6: Input parameters ($\psi^m$) for crossing decision model. The parameters are calculated for the current time instance when a gap starts unless mentioned otherwise.

| **Parameter** | **Description** |
|---|---|
| Veh. dist., $d_{pv}$ $[m]$ | Longitudinal distance between InterV and pedestrian |
| Veh. speed, $v_v$ $[m/s]$ | Speed of the InterV |
| Wait time, $w_p$ $[s]$ | Time elapsed since pedestrian started waiting |
| Gaze ratio, $\phi_g$ | Proportion of time pedestrian looked at the InterV in the previous second |
| Curb dist., $d_{curb}$ $[m]$ | Lateral distance between pedestrian & road edge |
| Crosswalk dist., $d_{CW}$ $[m]$ | Longitudinal distance between pedestrian & crosswalk |
| Ped. speed, $v_p$ $[m/s]$ | Average pedestrian speed in the previous second |
| Vehicle lane, $v_l$ | Boolean variable indicating whether the InterV is in the near lane or far lane with respect to the pedestrian |
| Vehicle direction, $v_d$ | Boolean variable indicating whether the InterV is approaching the pedestrian from the same direction as the pedestrian or the opposite direction |

We were interested in predicting pedestrian behavior from the perspective of the ego-vehicle (EgoV). Thus we assumed that at any given time, one of the vehicles in the scenario is an EgoV and predicted pedestrian trajectories from the perspective of that vehicle. The EgoV can either be the InterV (refer Figure 4.8) or an observer of the interaction between the pedestrian and the InterV (refer Figure 4.9). EgoVs are primarily interested in pedestrians who pose safety concerns during navigation. We defined an interaction region, $I$, by its radius $r_I$ (refer Figure 4.9) for the EgoV that excludes far-away pedestrians and pedestrians behind the vehicle. The interaction region is within the sensing range of the vehicle [167]. We assumed that the EgoV could accurately track the positions of the pedestrians and the other vehicles within the interaction region. The EgoV predicts pedestrians' behavior when they are within the interaction region. Because the MHP model was developed for single pedestrians, the implementation predicts the behaviors of each pedestrian within $I$ independently.

Predictions, by their very nature, are inherently uncertain. This model primarily focused on the uncertainties in pedestrian decision-making and actions (i.e., discrete state uncertainty). Pedestrians' trajectories can be significantly different based on their decisions, such as the decision to cross or wait to cross. The steps for real-time

trajectory prediction using the MHP model are given in Algorithm 2.

We used a Bayesian inference framework with two stages: *predict* and *update*. In the *predict* stage, whenever a pedestrian is detected within $I$, the EgoV initializes a prediction tracklet (e.g. $PT_0$ in Figure 4.9b) for that pedestrian with the current observed states of the pedestrian. These states include the position, velocity, heading, gaze, and the discrete state of the pedestrian. A *tracklet* is a part of the pedestrian trajectory that the pedestrian takes before reaching the next decision point. The discrete states are initialized based on the continuous state guard conditions defined in Section 4.5. The tracklet is propagated using the continuous dynamics model associated with the initialized discrete state until it reaches a decision point. We defined three types of decision points—(i) decision to attempt crossing or walk straight ahead without crossing, (ii) decision to accept the currently available gap or not, and (iii) decision on which direction to turn after crossing the street (refer Figure 4.9b). When the current tracklet reaches a decision point, that tracklet branches into two tracklets based on the associated probabilities of the future discrete states. There could be multiple decision points with the prediction horizon and the prediction branching into two at every decision point resulting in multiple possible trajectories. The number of possible trajectories, $\alpha$, is one more than the number of such decision points. Some of these decision points trigger discrete state transitions: approach to walkaway, approach to wait, wait to cross, etc. In contrast, others, such as the decision on the direction after crossing, result in multiple trajectories for the same "walkaway" discrete state.

The probability that an approaching pedestrian has the intent to cross, $p(C)$, is given by the crossing intent model (refer Section 4.5.1); the probability that the pedestrian will decide to cross, $p(GA)$, is given by the crossing decision model (refer Section 4.5.2) and we assumed equal probability for both turning directions after crossing. At the end of the prediction horizon $N$, the probability of the multimodal trajectories is given by the product of the probabilities of the constituent tracklets. The continuous state is predicted using the constant velocity model associated with the discrete state. The uncertainty in the continuous state predictions is expressed as Gaussian noise (refer Equation 4.7). A linear Kalman filter is used to predict and update the continuous states.

There is a delay between deciding to cross and the start of crossing [158]. We express this delay as $\tau_{cross}$ and sample it from an exponential distribution learned from each of the datasets. When a gap is accepted, the transition from wait to cross occurs when it is $\tau_{cross}$ seconds since the time gap was accepted (i.e., time delay is

**Algorithm 2** Real-time trajectory prediction using MHP model.

Initialize Tracklet $X_0, Q_0, \lambda_0^m, \psi_0^m$
**while** $t_i < T$ **do**
  **procedure** PREDICT
    **for** $k \leftarrow 1$ to $N$ **do** ▷ For entire prediction horizon
      **for** $j \leftarrow 1$ to $N_t$ **do** ▷ For all tracklets
        **if** Reached Decision Point **then**
          Predict decision probability given $\lambda_j^m, \psi_j^m$
          Create new tracklets (branches)
          Predict $Q_k$
        **end if**
        **if** Discrete State Transitioned ($Q_k \neq Q_{k-1}$) **then**
          Reset $X_k$ corresponding to $Q_k$
        **end if**
        Predict $X_{k+1} = f_{Q_k}(X_k, \mathcal{N}(0, W))$
        $j \leftarrow j + 1$
      **end for** ▷ end of all tracklets
      $k \leftarrow k + 1$
    **end for** ▷ end of prediction horizon
  **end procedure**
  **procedure** UPDATE
    $X_{i+1} \leftarrow$ observation $\hat{O}_{i+1}$
    $Q_{i+1} \leftarrow X_{i+1}, T(Q_{i+1}|Q_i)$
  **end procedure**
  $i \leftarrow i + 1$
**end while**

reached). Similarly, the pedestrian speed when starting to cross, $v_{start}$, is sampled from a Gaussian distribution, and the parameters are learned from the data. Because no new measurement is available within the prediction loop, the features $F_t$ within the prediction horizon are calculated from the predicted pedestrian positions. We assume that the gaze ratio of pedestrians remains the same for the entire prediction horizon as the most recent observation. Using predicted data instead of actual measurements for the features, the MHP model is able to predict pedestrian trajectories over long durations ($> 5\,s$). The parameters used in the model are tabulated in Table 4.7.

In addition to the tracklets formed by the decisions made by the pedestrians, we considered a tracklet with a non-zero probability that does not have a discrete state but always followed constant velocity dynamics, similar to the baseline constant velocity model. This ensured that the predictions from the MHP model were performance-bounded by the constant velocity predictions, i.e., the model would perform at least as well as the constant velocity model.

During the *update* stage, the continuous state predictions are updated based on the observations of the current time step. The discrete state probabilities are updated based on the guard conditions (refer Section 4.5) on the now updated continuous

states.

Table 4.7: MHP model parameters.

| **Parameter** | **Value** |
|---|---|
| Heading threshold, $\theta_h$ | $45^\circ$ |
| Decision zone length, $D$ | $3\,m$ |
| Interaction region radius, $r_I$ | $50\,m$ |
| Prediction horizon, $N$ | $0.2 - 6\,s$ |
| Sampling time, $\Delta t$ | $0.2\,s$ |
| Maximum pedestrian velocity, $v_{max}$ | $2.5\,m/s$ |
| Prediction envelope probability threshold, $\delta_k$ | 0.01 |

### 4.5.4 Datasets description

We developed and tested the MHP model on two urban datasets—Automated Vehicle Interaction in Virtual Reality (AVIVR) dataset and Intersection Drone dataset (inD) [37].

The inD dataset is a publicly available dataset collected using a drone and provides a bird's-eye view of a four-way unsignalized intersection in Germany. The dataset was collected at a sampling time of $0.05\,s$ and contains trajectory information of pedestrians, human-driven cars, and bicyclists (bicyclists, however, are not considered in this dissertation). The dataset contains 2100 pedestrian trajectories in total. We treated groups of pedestrians as multiple individual pedestrians, and their behaviors were predicted individually. We isolated scenes that had at least one vehicle. Gaze information was not directly available in this dataset, and we used pedestrian heading to approximate their gaze behavior.

Both datasets provide pedestrian and vehicle trajectories in absolute coordinates. The MHP model is run using these absolute coordinates. The model can be easily adapted for relative coordinates from the EgoV's frame of view. We downsampled both the datasets with a sampling time of $0.2\,s$ to maintain uniformity among the datasets and to align with previous works [38, 84, 168].

### 4.5.5 Baseline models

We compared the performance of the crossing intent and crossing decision models with standard logistic regression baselines, which is a common approach for modeling pedestrian behaviors [43, 169].

Collision avoidance systems typically use a constant velocity model for pedestrian trajectory prediction [159]. Thus, similar to [21], and [35], we compare our MHP model against a baseline model with pure constant velocity dynamics without any discrete states or switching dynamics. We also compared the MHP with our baseline UHP model developed in Section 4.3. However the UHP model conservatively assumes that all pedestrians approaching the crosswalks have the intent to cross and produces unimodal predictions (predicts only the most probable trajectory).

There are several differences in the two datasets we considered—AVs vs. human-driven cars, midblock vs. intersection, IVE vs. the real world, etc. Thus, we trained the crossing decision SVM model on each dataset separately and evaluated their performance separately.

### 4.5.6 Evaluation metrics

We evaluated the classification performance of the crossing intent and crossing decision models. Predictions were assigned to a class when the probability of the predicted class was greater than 0.5. We considered the standard classification metrics such as accuracy, precision, recall, and F1-score.

The real-time trajectory prediction performance was evaluated using the metrics of F1-score, ADE, and FDE described in Section 4.3.4. Additionally, we evaluated the MHP model performance with the probabilistic metrics of expectation of ground truth and the forward reachable set ratio.

- **Expectation of ground truth (EGT)**: expectation that the ground truth is covered by the prediction envelope for horizon $N$.
- **Forward reachable set ratio (FRSR)**: the ratio of the size of the prediction envelope to the full forward reachable set.

Consider the ground truth position at time step $i$ to be given by $(x_i, y_i)$. Similarly, the predicted position corresponding to the future $j$ is given by $(\hat{x}_i^j, \hat{y}_i^j)$, where $j \in [1, \alpha]$ and $\alpha$ is the number of possible futures. The ADE and FDE corresponding to the future $j$ are thereby given as,

$$ADE_j = \frac{\sum_{i=1}^{N} \|(x_i, y_i) - (\hat{x}_i^j, \hat{y}_i^j)\|^2}{N}, \tag{4.8}$$

$$FDE_j = \|(x_N, y_N) - (\hat{x}_N^j, \hat{y}_N^j)\|^2. \tag{4.9}$$

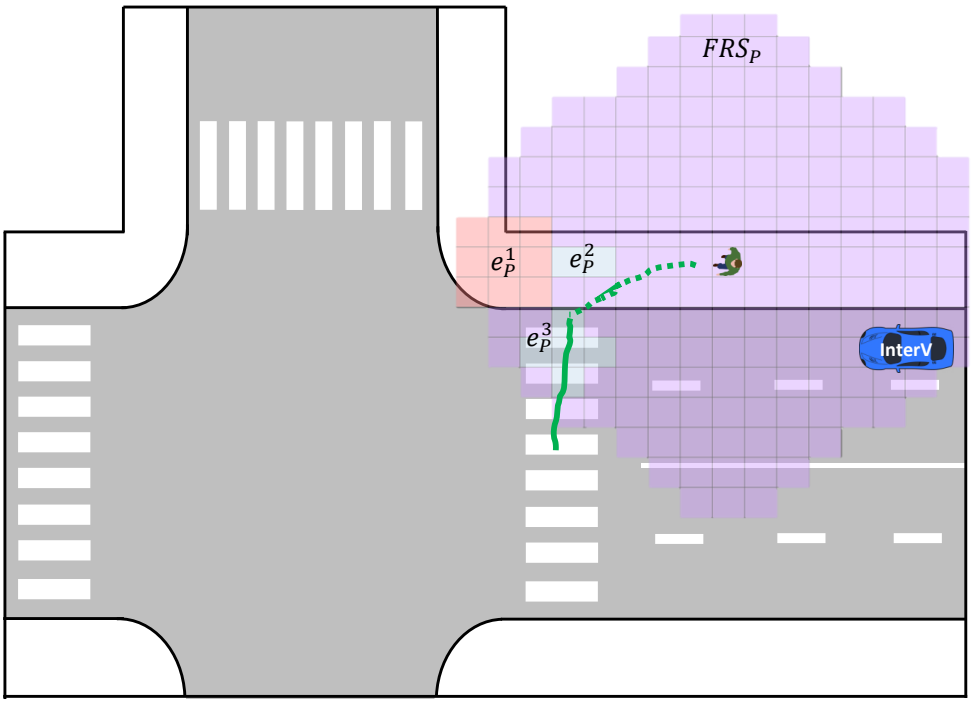



Figure 4.10: Consider an example scenario of trajectory prediction with a pedestrian approaching a crosswalk. $e_p^1\, e_p^2$ and $e_p^3$ are prediction envelopes at prediction time step $t_P$. $e_p^1$ is the prediction envelope of the constant velocity model and $e_p^2, e_p^3$ are the prediction envelopes corresponding to two possible future behaviors—waiting by the crosswalk and crossing—identified by the MHP model. $FRS_P$ is the forward reachable set at $t_P$ assuming the pedestrian could have walked in any direction with a maximum speed of $2.5\,m/s$. The green line indicates observed ground truth trajectory up to time $t_P$ and the solid orange line indicates the ground truth trajectory after $t_P$. The constant velocity predicts envelope $e_p^1$ based on the initial heading of the pedestrian. However, the pedestrian turns to cross at the crosswalk, which is captured by the prediction envelope $e_p^3$ of the MHP model.

The environment where the pedestrians are walking (roads, sidewalks, crosswalks, etc.,) is discretized into $0.2\times0.2\,m$ grids (approximately the space occupied by a standing pedestrian). The MHP model, at every time step, predicts all possible trajectories where the probabilities of the trajectories are calculated from the probabilities of the decisions made. The continuous states are estimated using a Kalman filter with a constant velocity prediction model. For each trajectory, the position at each time step is distributed as Gaussian noise $W$ (refer Equation 4.7). The noise is propagated for the entire prediction horizon $N$ through the Kalman filter. As shown in Figure 4.10, each trajectory had an associated prediction envelope ($e_i^j$) given by the variance (Gaussian noise) propagated by the Kalman filter for that trajectory at time step

$i$. All the individual prediction envelopes combined to form the prediction envelope for that time step, $E_i = \{e_i^1, e_i^2, ...., e_i^\alpha\}$. Only the grids with a minimum probability threshold of $\delta_k$, similar to [170], were considered to be part of the prediction envelope.

EGT metric is the average of a Boolean variable, $\nu$, at every prediction time step averaged over the entire prediction horizon. $\nu$ is '1' if the ground truth position at a particular prediction time step is within the prediction envelope of that time step and '0' otherwise. EGT is thereby defined as,

$$EGT = \frac{\sum_{i=1}^{N} \nu_i}{N}, \tag{4.10}$$

$$\text{where} \quad \nu_i = \begin{cases} 1 & \text{if } (x_i, y_i) \in E_i \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

For computing the FRSR metric, the fully forward reachable set, $FRS$, of the pedestrian is first calculated. The maximum velocity of the pedestrian is assumed to be $2.5\,m/s$ (estimated from the inD dataset) and a circular $FRS$ is calculated (refer Figure 4.10). FRSR at the end of the prediction horizon $N$ is given as,

$$FRSR = \frac{FRS_N}{E_N}. \tag{4.12}$$

### 4.5.7 Results and discussion

In Sections 4.5.7.1 and 4.5.7.2, we report the performance of the crossing intent and crossing decision models respectively. In Section 4.5.7.3, we report the trajectory prediction results of the MHP model, and finally in Section 4.5.7.4, we present the computation performance of the MHP model.

#### 4.5.7.1 Crossing intent

The crossing intent model was evaluated only on the inD dataset because all pedestrians crossed the street in the AVIVR dataset. The features ($\lambda_m$) for the crossing intent model had an observation window of $3\,s$ and were compiled every $1\,s$ (rolling window) for the duration when a pedestrian was approaching an unsignalized crosswalk. Each observation window had an associated class—'intend to cross', and 'do not intend to cross'—where a pedestrian is considered to intend to cross if they started crossing at any time during the time they were observed by the EgoV (irrespective of the actions or even the presence of the InterV). Features ($\lambda_m$) and their associated class were calculated from the observations of 2100 pedestrians when they were ap-

proaching a crosswalk—1470 of them with the intent to cross and 630 without the intent to cross—resulting in the cross intent dataset.

The dataset was split into the train (80%) and test (20%) sets and bootstrapped the training set to balance the training data for the two classes. The training data were used to develop the model, and the developed model was evaluated on the unseen testing data. After testing the performance of various kernel functions (ranging from the simple kernels to more complicated kernels) for the SVM classifier—linear, quadratic, cubic, Gaussian—we chose the Gaussian Kernel for the SVM model since it had the highest performance in terms of F1-score. For an observation duration of $3\,s$, the model performance is shown in Table 4.8. The SVM model performed better, with higher accuracy and F1-score, than a baseline logistic regression model in predicting the crossing intent of approaching pedestrians and was used for pedestrian trajectory prediction.

Table 4.8: Crossing intent model performance for the inD dataset for identifying Cross (C) and Not Cross (NC) intents.

| **Model** | **Accuracy** | **Precision (C/NC)** | **Recall (C/NC)** | **F1-Score (C/NC)** |
|---|---|---|---|---|
| Logistic Regression | 0.80 | 0.89/0.36 | 0.88/0.40 | 0.88/0.38 |
| Support Vector Machine | **0.89** | **0.95/0.59** | **0.92/0.69** | **0.93/0.63** |

Table 4.9 shows the importance of the various parameters based on the SVM model performance [160]. The features are arranged in descending order of importance. It can be noted that removing even one parameter drastically affects the model performance (some more than others), indicating the importance of all eight parameters for crossing intent prediction. The lateral distance of the pedestrian with the curb had the most influence on model performance. This is probably because pedestrians who do not intend on crossing do not walk close to the curb.

The pedestrians' intent to cross was accurately predicted as early as 4.6 $m$ and on average 2.5 $m$ from the crosswalk. The intent to not cross was predicted as early as 7.4 $m$ and on average 4.4 $m$ from the crosswalk. These early intent predictions should aid AVs in identifying which pedestrians to focus future predictions on and generally improve the prediction performance as demonstrated in Section 4.5.7.3.

#### 4.5.7.2 crossing decision

We used the crossing decision SVM model for predicting the crossing decision. Instances of traffic gaps were identified when a pedestrian was inside the decision zone

Table 4.9: Crossing Intent model feature ranking based on the performance in identifying cross intent (C).

| **Feature removed** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|
| Curb dist. | 0.70 | 0.65 | 0.86 | 0.76 |
| Ped. speed | 0.72 | 0.67 | 0.86 | 0.78 |
| Veh. speed | 0.73 | 0.69 | 0.85 | 0.77 |
| Gaze ratio | 0.75 | 0.70 | 0.88 | 0.78 |
| Veh. dist | 0.79 | 0.70 | 0.89 | 0.80 |
| Veh. Lane | 0.79 | 0.75 | 0.86 | 0.80 |
| Crosswalk dist. | 0.80 | 0.74 | 0.88 | 0.81 |
| Veh. direction | 0.82 | 0.78 | 0.89 | 0.84 |

and when a traffic gap had just started. Every such instance set had an associated class—'accept gap', and 'reject gap'—where a pedestrian is considered to accept the gap if they started crossing at any time during the time the were interacting with the InterV corresponding to this traffic gap and reject the gap otherwise. For model training and testing, we calculated the features ($\psi^m$) and their associated class at these instances—508 accepted gaps and 1,195 rejected gaps for the AVIVR dataset and 601 accepted gaps and 2,926 rejected gaps for the inD dataset—resulting in the gap datasets.

We split the data into training (80%) and testing (20%) sets for both the datasets and bootstrapped the training data to balance the different classes. After testing the performance of various kernel functions (ranging from the simple kernels to more complicated kernels) for the SVM classifier—linear, quadratic, cubic, Gaussian—we chose the Gaussian Kernel for the SVM model since it had the highest performance in terms of F1-score. The model that was developed using the training data was then evaluated on the unseen testing data. We compared the model with a baseline logistic regression model, similar to [43]. As shown in Table 4.10, the SVM model performed better, with higher accuracy and F1-score than the baseline, and was used for pedestrian trajectory prediction.

The SVM model showed similar performance in both AVIVR and inD datasets. In the inD dataset, the scene was a four-way intersection with multiple vehicles on two-way streets and multiple pedestrians interacting. On the other hand, the AVIVR dataset scene was a midblock crosswalk with a single pedestrian and AVs approaching from a single direction on a one-way street. Nevertheless, the model showed good performance in both the unsignalized midblock and unsignalized intersection scenarios, indicating the model's applicability to various situations.

Table 4.10: Crossing decision model performance for AVIVR and inD datasets for accepting a gap.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Logistic Regression (AVIVR) | 0.79 | 0.71 | 0.52 | 0.60 |
| Support Vector Machine (AVIVR) | **0.88** | **0.75** | **0.73** | **0.74** |
| Logistic Regression (inD) | 0.74 | 0.67 | 0.63 | 0.65 |
| Support Vector Machine (inD) | **0.84** | **0.79** | **0.70** | **0.75** |

#### 4.5.7.3 Trajectory prediction

In this section, we discuss the evaluation of the MHP model for predicting pedestrian trajectories using the steps discussed in Algorithm 2. The MHP model was evaluated on both the AVIVR and the inD datasets. For deterministic error metrics, we compared two types of trajectories from the MHP model—the most probable prediction and the best prediction—with the ground truth trajectory. In the AVIVR dataset, however, the MHP model had a fixed probability of crossing intent, i.e., $p(C) = 1$, because all pedestrians in the AVIVR dataset crossed the street. In the AVIVR dataset, the most probable prediction of the MHP model would be the same as the prediction from the baseline hybrid model because all pedestrians had the intent to cross and the crossing decision was the only variable.

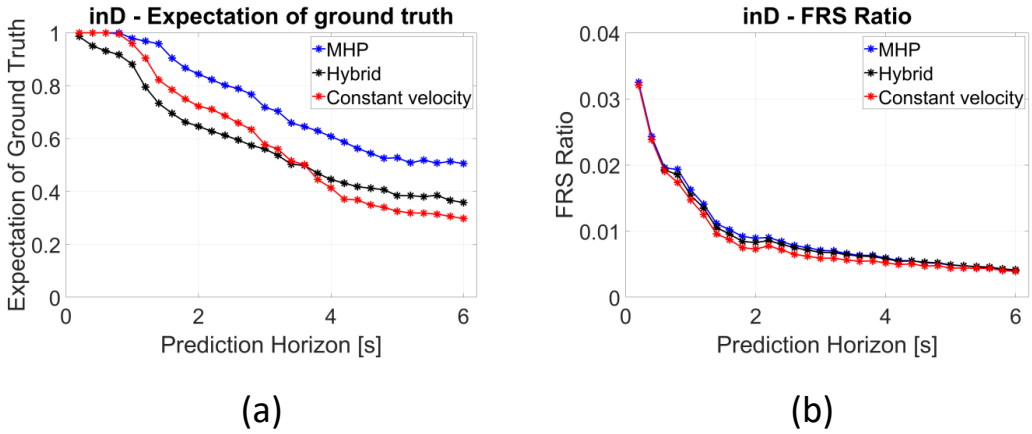



Figure 4.11: Comparison of (a) expectation of ground truth and (b) FRS ratio of the MHP model with the baseline models. The ground truth trajectory is more likely to be captured in the MHP model, with a larger prediction envelope at lower prediction horizons as seen from the slightly higher FRS ratio for the MHP model.

From Figure 4.12a and 4.12b, it can be seen that for the AVIVR dataset, the baseline hybrid and MHP models performed better than the constant velocity baseline. The predictions of the baseline hybrid and MHP models were the same because all

pedestrians had the intent to cross in this dataset. For the inD dataset, it can be observed that though the most probable prediction was not better than the baselines, the best prediction of the MHP model performed better than both the hybrid model and constant velocity model baselines (refer Figure 4.12c and 4.12d) for the entire prediction horizon, especially at longer prediction horizons. This shows that at least one trajectory among the multiple trajectories performed well. The difference in performance between the MHP model and the baselines increased with the prediction horizon.

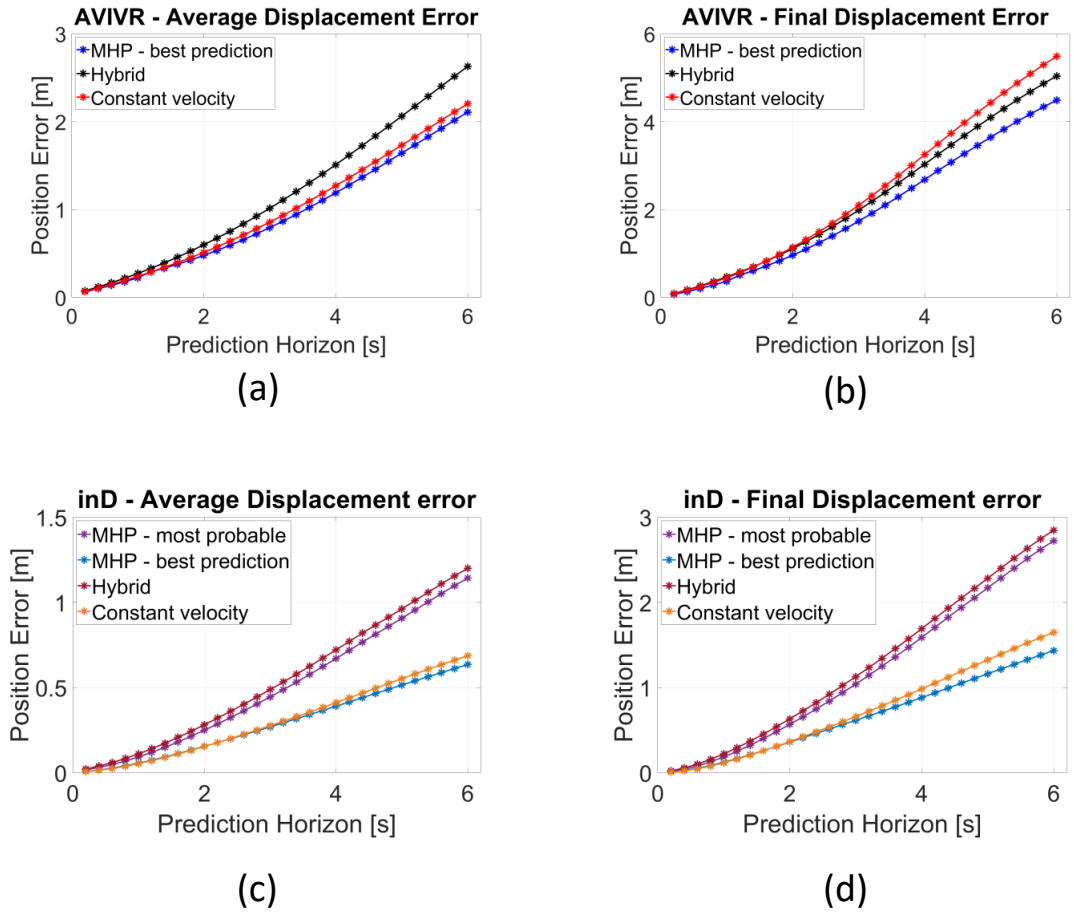



Figure 4.12: Pedestrian trajectory prediction comparison in AVIVR dataset at different prediction horizons for (a) average displacement error metric, and (b) final displacement error metric. The hybrid model has lower error than the constant velocity baseline. Pedestrian trajectory prediction comparison in inD dataset at different prediction horizons for (c) average displacement error metric, and (d) final displacement error metric. The most probable prediction from the MHP model is not necessarily better than the baselines but the best predicted trajectory (one closest to ground truth trajectory) from the MHP model has lower error than both the hybrid model and constant velocity baselines.

The performance of the MHP model on the AVIVR and inD datasets exhibited its capability (i) for predicting multiple possible trajectories—at least one of which at any time is closer to the actual ground truth trajectory than the baseline predictions, (ii) for prediction under different interaction situations—unsignalized midblocks and intersections, and (iii) for situations containing interactions with both human-driven and automated vehicles. However, one limitation of the MHP model is that because of the vast differences in the two types of datasets, the individual crossing intent and crossing decision SVM models had to be independently trained on each dataset.

We evaluated the performance of the MHP model on the inD dataset through probabilistic metrics. Figure 4.11 shows the expectation of ground truth and FRS ratio for various prediction horizons. The expectation of ground truth gives the likelihood that the ground truth trajectory is in the set of multimodal predicted envelopes. It can be observed that the probability was higher at lower time steps and decreased at higher time steps. The MHP model was more likely to capture the ground truth than the baselines for all time horizons. This normally occurred with an increase in the FRS ratio of the MHP compared to the baselines because of the multiple possible prediction envelopes. A high FRS ratio indicates more conservative behavior because the prediction envelopes include a higher proportion of possible behaviors. The FRS ratio of the MHP model was only slightly higher than the baselines and further converged as the prediction horizon increased. Thus MHP model better captured the ground truth without being overly conservative.

The multimodal nature of pedestrian behavior results in uncertain pedestrian behaviors. The MHP model efficiently predicted multimodal trajectories for a long term. Unlike existing studies [27, 97], the multimodality was not arbitrarily conditioned on a latent variable, but instead was conditioned on the decision-making process of the pedestrian, which is grounded in actual pedestrian behavior and thereby easier to comprehend. It can be observed that the EGT metric was high initially and then decreased with an increase in the prediction horizon. Even at the end of $6\,s$, the MHP model was 50% likely to capture the ground truth, which is much higher than the baselines.

#### 4.5.7.4 Computational Performance

The experiments were run on Intel Core i7-7700K, 16 GB RAM, and Nvidia GTX 1080 Ti. On average, the MHP algorithm took 0.021 $s$ to run predictions for the entire prediction horizon of up to $6\,s$. The SVM models were inferred only during the decision points, whereas during other time instances, the MHP model ran simple

constant velocity models. This enabled the model to run for long-term and multiple predictions independently for each pedestrian in the scene.

## 4.6 Chapter summary

The contribution presented in this chapter is the development of an intuitive and computationally efficient pedestrian behavior model that is suitable for long-term multimodal predictions and has been published in [35, 36]. We developed two models of pedestrian behavior. The UHP model demonstrated the applicability of the hybrid automata framework for pedestrian trajectory prediction. The MHP model extended the framework to predict multimodal pedestrian trajectories. The models were validated in both VR and real-world inD datasets. We developed intuitive and computationally efficient models of pedestrian behavior based on hybrid automata theory that are suitable for predicting long-term multimodal pedestrian trajectories in urban environments and found the models to perform better than the baseline models.

The models developed so far focused on the interaction of a single pedestrian with an interacting vehicle (InterV). In the real world, the behavior of pedestrians may also be affected by the behavior of neighboring pedestrians. In Chapter V, we extend the hybrid automata framework to include the interactions between pedestrians.

# CHAPTER V

# Multiple Pedestrian Interaction Model

## 5.1 Overview of chapter

Pedestrian behavior is not only dependent on the interacting vehicle behavior but also on the behaviors of the nearby pedestrians [16]. Current pedestrian models incorporate interactions with other pedestrians primarily to avoid collisions, and neglect the effect of nearby pedestrians on the decision-making process of the pedestrian of interest (POI). Further, these models incorporate all possible interactions between multiple pedestrians and multiple vehicles in a data-driven framework that lacks intuition and increases problem complexity [28, 39]. The contribution in this chapter is the extension of the pedestrian prediction framework from Chapter IV to incorporate their interaction with neighboring pedestrians and an approaching vehicle. The extended model was validated on the diverse and large-scale nuScenes dataset containing real-world interactions between pedestrians and vehicles [40]. We also identified a limiting interaction region for pedestrians beyond which pedestrian behavior is not significantly affected by other pedestrians.

In urban scenarios, multiple pedestrians typically cross together. In such situations, the actions of one or more pedestrians can affect the decisions of the other pedestrians. A pedestrian might be more comfortable crossing a street if there are people already crossing the street (refer Figure 5.1). Further, pedestrians crossing as a group can have different behaviors in terms of gap acceptance [16].

In this chapter, we develop an intuitive hybrid automata model called interacting hybrid pedestrian (IHP) for long-term multimodal pedestrian trajectory prediction when interacting with AVs and other pedestrians. Results demonstrate the applicability of the IHP model for long-term ($> 5\,s$) pedestrian trajectory prediction at unsignalized crosswalks. We compared the IHP with the state-of-the-art Trajec-

tron++ model [28] and found the IHP model to have a similar multimodal performance to the Trajectron++ model at a lower computational cost.

Figure 5.1: Multiple pedestrians crossing a street. Crossing decisions of pedestrians can be influenced by the presence of nearby pedestrians. For example, the pedestrian on the right is more likely to cross the street when they see pedestrians already crossing. Picture from nuscenes dataset [40]

## 5.2 Pedestrian interaction in urban scenarios

We still focus our efforts on pedestrians around unsignalized crosswalks. As discussed in Section 5.1, pedestrian decisions can be influenced by the behavior of nearby pedestrians. The POI can also be a part of a group (friends, colleagues, etc.) that intend to travel to a destination together. Pedestrian group behavior has been observed to be significantly different from individual pedestrian behavior [171]. Group size is one of the most influential factors affecting pedestrian behavior [16]. When crossing as a group, pedestrians tend to be more careless, and pay less attention at crosswalks and often accept shorter gaps between the vehicles to cross [171, 172] or do not look for approaching traffic [48]. Thus the behaviors of nearby pedestrians need to be considered for better predicting the behavior of POI. However, not all the nearby pedestrians will have equal influence on the POI. The closer pedestrians have more influence on the behavior of the POI than the farther pedestrians.

## 5.3 Interacting hybrid pedestrian model

In this section, we discuss the development of the interacting hybrid pedestrian (IHP) behavior model. The model is an extension of the MHP model discussed in Section 4.5. Similar to MHP, the IHP model also predicts multimodal trajectories of pedestrians and utilizes a hybrid automata framework.

The IHP model applies to both unsignalized midblocks and intersections and considers vehicles moving in any direction. Additionally, the model also includes situations where pedestrian-vehicle interactions occur on multi-lane roads. The model predicts the behavior of the pedestrian of interest (POI). As in the MHP model, the observations of the traffic are through the ego-vehicle termed as $EgoV$ and the vehicle closest to the POI and approaching the POI is called the interacting vehicle ($InterV$). The InterV may be the EgoV or another vehicle that the EgoV observes interacting with the POI.

We do not assume that the interactions between POI and the InterV are isolated. Instead, the neighboring pedestrians can influence the interactions within a region surrounding the POI called the interaction region (refer Figure 5.3a). The interaction region is defined by two parameters, the interaction distance $h$ and interaction angle $\theta$. The human's viewing frustum inspires this conical interaction region as pedestrians are highly likely to interact with the other pedestrians who are in their field of view [173].

### 5.3.1 Hybrid automata model

Given a history of observations for $M$ time steps, denoted by $\hat{O}_{t-M:t}$, we would like to predict the pedestrians' trajectories $N$ time steps into the future given by, $\tilde{X}_{t+1:t+N} \in \mathbb{R}^2$. The observations for the $M$ time steps $\hat{O}_{t-M:t}$ include the trajectory of the pedestrians (POI and neighbouring pedestrians) $\hat{X}_{t-M:t} \in \mathbb{R}^2$, the gaze of the pedestrians $\hat{G}_{t-M:t} \in \mathbb{R}^3$, and the trajectory of the interacting vehicle $\hat{X}^i_{t-M:t} \in \mathbb{R}^2$. Other information such as pedestrian and vehicle velocities and relevant distances are calculated from the observed pedestrian and vehicle trajectories (refer Tables 5.1 and 5.2).

Like MHP, the IHP is generalizable to midblocks and intersections and can model all possible pedestrian action sequences. We focused only on rational pedestrians who either cross at a crosswalk without jaywalking or do not intend to cross. Similar to the MHP model (refer Section 4.5), the continuous dynamics is represented using a constant velocity model with Cartesian position and velocity as the states, $X =$

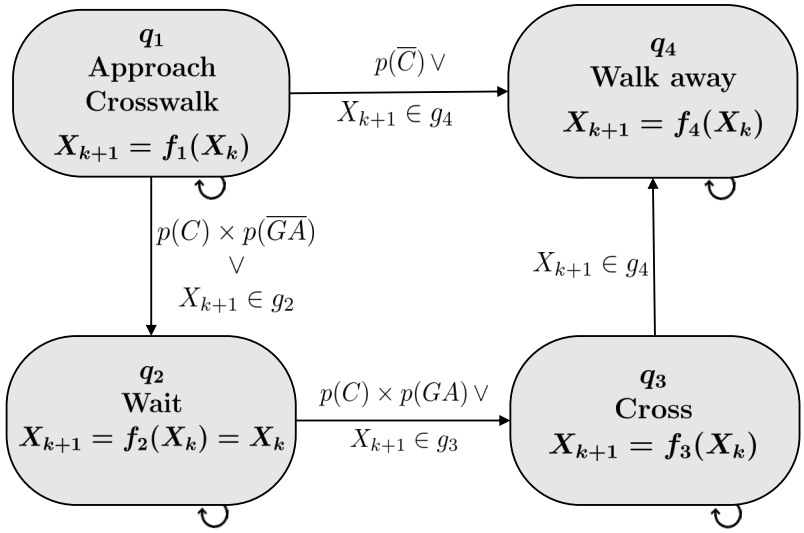



Figure 5.2: IHP, a probabilistic hybrid automaton model of a pedestrian. The transitions between states $q_1$ and $q_2$ and between $q_2$ and $q_3$ are determined by the predictions from the crossing intent model and crossing decision model. This framework is similar to the MHP model discussed in Section 4.5. The difference lies in the probabilistic model of crossing decision which incorporates additional features of pedestrian-pedestrian interaction.

$(x, y, v_x, v_y)$ and zero-mean Gaussian process noise (refer Equation 4.7).

The hybrid automaton has four discrete states—approach crosswalk, wait, cross, and walk away—each with an associated continuous motion. The transitions between the discrete states are non-deterministic and probabilistic and depend on (i) the intent of the pedestrian to cross the street, (ii) the gap acceptance behavior of the pedestrians, and (iii) the guard conditions on the continuous states of the pedestrian that differentiate the discrete states. We model pedestrians' crossing intent (refer Section 5.3.3) and crossing decision (refer Section 5.3.4) separately and incorporate these two models into the hybrid automaton framework.

The proposed IHP model (shown in Figure 5.2) is formally defined as a tuple $\langle X, Q, f, G, T, R \rangle$, where

- $X = (x, y, v_x, v_y) \in \mathbb{R}^4$, are the continuous pedestrian states of Cartesian positions $(x, y)$ and velocities $(v_x, v_y)$.
- $Q \in \{q_1, q_2, q_3, q_4\}$, are the discrete states of approach, wait, cross, or walkaway respectively.
- $f : Q \times X \to \mathbb{R}^4$, represents the continuous dynamics.

- $G \in \{g_1, g_2, g_3, g_4\}$ is a set of guard conditions, where
  - $g_1 = \{X \mid v_x \neq 0 \wedge \theta \leq \theta_h\}$
  - $g_2 = \{X \mid v_x = 0 \wedge v_y = 0\}$
  - $g_3 = \{X \mid v_y \neq 0\}$
  - $g_4 = \{X \mid v_x \neq 0 \wedge \theta > \theta_h\}$,

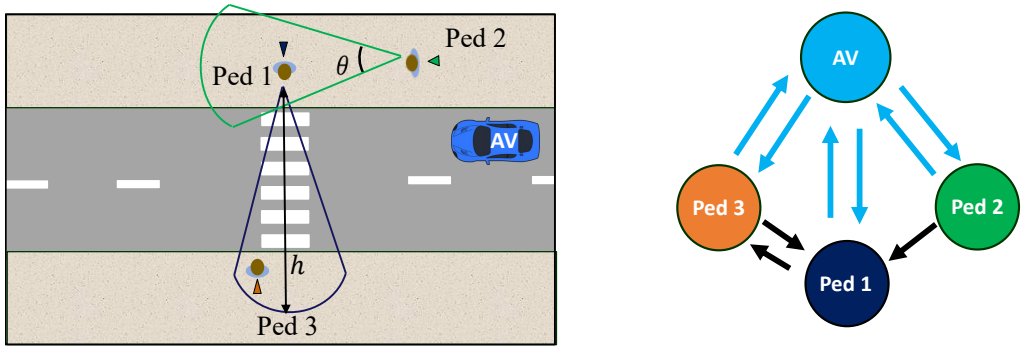
  where $\theta$ is the angle difference between pedestrian heading angle and the angle between pedestrian and the crosswalk of interest (refer Figure 4.8) and $\theta_h$ is a threshold value for heading towards the crosswalk of interest.

- $T$ is the discrete state transition function given as,
  - $T(Q_{k+1}|Q_k) = \begin{cases} q_1 & \text{if } X_{k+1} \in g_1 \\ q_2 & \text{if } X_{k+1} \in g_2 \vee p(C) \times p(\overline{GA}) \\ q_3 & \text{if } X_{k+1} \in g_3 \vee p(C) \times p(GA) \\ q_4 & \text{if } X_{k+1} \in g_4 \vee p(\overline{C}), \end{cases}$

  where $p(C)$ is the probability of the pedestrian having the intent to cross and $p(GA)$ is the probability of the pedestrian to accept a gap and cross the street.

- $R : X \to X$ is a reset map of the continuous velocity and heading states after the discrete state transitions.

Similar to the MHP model, in the IHP model, the transition function depends on both the probability of the pedestrian having the intent to cross and the probability of accepting a gap. Likewise, the IHP model is more generally applicable to an intersection scenario with multiple crosswalks. So, identifying the crosswalk the pedestrian is approaching is critical as the guard and reset conditions depend on the crosswalk location and road orientation. We describe pedestrian discrete state transition from wait to cross (i.e., crossing decision) through their gap acceptance behavior. We develop a model that outputs the probability of accepting a traffic gap. The primary difference between the IHP and the MHP models is the crossing decision model. In the IHP model, additional features are included in the crossing decision model to incorporate the interactions with nearby pedestrians.




(a) (b)

Figure 5.3: **(a)** A typical situation with three pedestrians with the intention to cross and one interacting AV; pedestrian 1 can affect the behavior of pedestrian 2 but the inverse is not true. **(b)** A graph representation of the scenario. The blue arrows indicate pedestrian-AV interactions and the black arrows indicate pedestrian-pedestrian interactions.

### 5.3.2 Assumptions for the IHP model

In addition to the assumptions discussed in Section 4.3 (except for assumption IV.5), we assume the following.

**Assumption V.1.** *The nearby pedestrians have weighted influence on the POI and the weight is inversely proportional to the distance between the pedestrians.*

**Assumption V.2.** *The interaction between a pair of agents (pedestrians or InterV) is asymmetrical.*

For example, as shown in Figure 5.3a, pedestrian 2 may be interacting with pedestrian 1 but pedestrian 1 need not necessarily also interact with pedestrian 2.

**Assumption V.3.** *Pedestrians interact with neighboring pedestrians only if they are within a specified interaction region, $I$.*

We define an interaction region $I$ for pedestrians. Only pedestrians within this interaction region interact with the POI as shown in Figure 5.3a. The interaction region is defined by a maximum interaction distance $h$ and a limiting interaction angle $\theta$. Unlike prior studies that arbitrarily fix the interaction region [28, 39], we evaluate the performance of the crossing decision model for several different combinations of the interaction region parameters to identify the limiting interaction region (refer Section 5.3.6).

Table 5.1: Input parameters ($\lambda^m$) for crossing intent model calculated for an observation window of $3\,s$.

| Parameter | Description |
|---|---|
| Gaze ratio, $\phi_g$ | Proportion of time pedestrian gazed at InterV |
| Mean curb dist., $\bar{d}_{curb}$ $[m]$ | Average lateral distance between pedestrian & road edge |
| Mean crosswalk dist. $\bar{d}_{CW}$ $[m]$ | Average longitudinal distance between pedestrian & crosswalk |
| Mean ped. vel., $\bar{v}_p$ $[m/s]$ | Average pedestrian speed |
| Std. curb dist., $\tilde{d}_{curb}$ $[m]$ | Standard deviation of curb distance |
| Std. crosswalk dist. $\tilde{d}_{CW}$ $[m]$ | Standard deviation of longitudinal distance between pedestrian & crosswalk |
| Std. ped. vel., $\tilde{v}_p$ $[m/s]$ | Standard deviation of pedestrian speed |

As discussed in Section 2.4, recent works [28, 39, 174] have expressed the interaction between traffic agents either (i) by pooling abstracted hidden states of nearby pedestrians or (ii) as a spatial graph. Pooling uses hidden states of the pedestrians, expecting such states to capture the pedestrians' motion properties. However, trajectories are continuous and do not have such abstract "hidden states" [175]. The physical meaning of these abstracted hidden states is difficult to interpret, and using these "states" to represent the motion properties is indirect and non-intuitive. The graph-based approach captures the relationships between agents directly and enables the use of the physical states of the pedestrians, which is more intuitive.

We also express the scene as a spatial graph (refer Figure 5.3b) $G = \{V, E\}$, where $V$ represents the set of nodes, i.e., the individual agents, and $E$ represents the set of edges or interactions between the agents.

### 5.3.3 Crossing intent model

Pedestrians walking on the street may not always intend to cross the street. We developed a model to predict pedestrians' probability to cross a street while approaching the crosswalk. We modeled this as a probabilistic classification problem using a support vector machine (SVM) [154]. Pedestrians intend to cross a street at a particular crosswalk based on their goal location. When they intend to cross at this crosswalk, they are likely to walk closer to the curb and slow down as they are approaching the crosswalk. This behavior would be independent of other factors such as the behaviors of the vehicles and the other pedestrians. Thus we use pedestrian-related features to model their crossing intent.

We observed pedestrian trajectory and gaze for an observation window of $3\,s$ (we tested observation durations of $1\,s$, $2\,s$, $3\,s$, and $4\,s$ and $3\,s$ gave the best perfor-

mance). The input parameters for the model are the same as the parameters used in the crossing intent model of the MHP model (denoted by $\lambda^m$) and are detailed in Table 5.1. Crossing intent influences the discrete state transitions of *approach to wait* and *approach to walkaway*.

### 5.3.4 Crossing decision model

We describe the pedestrian's discrete state transition of *wait to cross* (i.e., crossing decision) through their gap acceptance behavior [43]. We develop a model that outputs the probability of accepting a traffic gap. We model the decision to cross using a Support Vector Machine (SVM) classifier and obtain probabilistic outputs following the method in [154].

We consider neighboring pedestrians within the interaction region of the POI as affecting the behavior of the POI. We consider that the effect of the pedestrians, given by $w_i$, depends on their distance to the POI, with the closer pedestrians having a stronger influence on the behavior of the POI. Similar to [174], we consider the effect of $N_p$ neighboring pedestrians on the behavior of the POI to be linear and can thus be aggregated by summing. This aggregation results in one equivalent interacting pedestrian $P_{eq}$ that has the same effect on the crossing decision of the POI as the $N_p$ nearby pedestrians. The aggregate interaction distance is the weighted sum of the distances to all interacting pedestrians from the POI.

The weights (effects) of the neighboring pedestrians are inversely proportional to their distance from the POI and are given as,

$$w_i = \frac{c}{{d_i}^2}. \tag{5.1}$$

where $c$ is a scaling parameter.

The aggregate interaction distance is given as,

$$d_a = \sum_{i=1}^{N_p} w_i' d_i, \tag{5.2}$$

where $N_p$ is number of interacting pedestrians and $w_i'$ is the normalized weight (relative to $w_i$) of the interacting pedestrians.

We include this aggregate interaction distance in the feature set for model development. We consider only the closest pedestrian action as influencing the action transition of the POI and include the discrete state of the closest pedestrian in the feature set.

Pedestrians can be walking by themselves or a part of a group. A group of pedestrians is expected to have different road crossing behaviors than individual pedestrians. For example, groups of pedestrians are more willing to take risks by accepting smaller traffic gaps [16]. A neighboring pedestrian is a part of a group with the POI if the distance between them was within $1\,m$ for at least 75% of the duration both pedestrians were observed in the scene [176]. Several such neighboring pedestrians can form a group with the POI. We include the group size parameter in the feature set (refer in Table 5.2) to model the crossing decision.

Table 5.2: Input parameters ($\psi^{ip}$) for crossing decision model. The parameters are calculated at the time instance when a gap starts unless mentioned otherwise.

| **Parameter** | **Description** |
|---|---|
| AV distance $[m]$ | Longitudinal distance between AV and pedestrian |
| AV speed $[m/s]$ | Speed of the AV |
| Curb distance $[m]$ | Lateral distance between pedestrian & road edge |
| Crosswalk distance $[m]$ | Longitudinal distance between pedestrian & crosswalk |
| Ped. speed $[m/s]$ | Average pedestrian speed in the previous second |
| Vehicle lane, $v_l$ | Boolean variable indicating whether the InterV is in the near lane or far lane with respect to the pedestrian |
| Vehicle direction, $v_d$ | Boolean variable indicating whether the InterV is approaching the pedestrian from the same direction as the pedestrian or the opposite direction |
| Interaction distance, $d_a$ | Aggregate interaction distance of all interacting pedestrians |
| Discrete state, $Q_{InterP}$ | Discrete state of closest interacting pedestrian |
| Group size, $N_p$ | Number of interacting pedestrians who form a group |

The input features ($\psi^{ip}$) to the SVM model include the previous set of features used in the MHP model ($\psi^{m}$, refer Table 4.6) and an additional set of features capturing the effects of the neighboring interacting pedestrians. These additional features are aggregate interaction distance, discrete state of closest interacting pedestrian (InterP),

and the group size. All the features are detailed in Table 5.2.

### 5.3.5 IHP model implementation for real-time trajectory prediction

A prediction framework similar to the one discussed in Section 4.5.3 is used. The framework uses the individual models of crossing intent and crossing decision to predict the multimodal pedestrian behaviors. The framework requires the real-time calculation of curb distance and crosswalk distance which requires map information. Vectorized map information is fed to the prediction algorithm, which contains the polygon vertices of two types of map regions, namely, crosswalks and road lanes.

There is a delay between deciding to cross and the start of crossing [158]. We express this delay as $t_{cross}$ and sample it from an exponential distribution learned from the dataset. The transition from *wait to cross* occurs when it is $t_{cross}$ seconds since the time gap was accepted (i.e., the time delay is reached). Similarly, the pedestrian speed when starting to cross, $v_{start}$, is sampled from a Gaussian distribution learned from the data. During the prediction stage, we assume constant velocity dynamics for the AVs. When no new measurement is available (i.e., within the prediction horizon), the features $\psi^{ip}$ are calculated from the predicted pedestrian and AV positions. We assume that the gaze ratio of pedestrians remains the same for the entire prediction horizon as the most recent observation. Using the predicted features enables the model to predict pedestrian trajectories for long durations.

We are interested in predicting pedestrian behavior from the perspective of the ego-vehicle (EgoV). The EgoV can either be the InterV (refer Figure 4.8) or an observer of the interaction between the pedestrian and the InterV (refer Figure 4.9). We assume that the EgoV could accurately track the positions of the pedestrians and the other vehicles within its sensing zone.

Predictions, by their very nature, are inherently uncertain. We primarily focused on the uncertainties in pedestrian decision-making and actions (i.e., discrete state uncertainty) in this work. Pedestrians' trajectories can be significantly different based on their decisions, such as whether to cross or wait to cross. The steps for real-time trajectory prediction using the IHP model are given in Algorithm 3. The real-time prediction follows the same steps as that in Section 4.5.3.

We used a Bayesian inference framework with two stages: *predict* and *update*. In the *predict* stage, whenever a pedestrian is detected within $I$, the EgoV initializes a prediction tracklet (e.g. $PT_0$ in Figure 4.9b) for that pedestrian with the current observed states of the pedestrian. These states include the position, velocity, heading, gaze, and the discrete state of the pedestrian. The discrete states are initialized

based on the continuous state guard conditions defined in Section 5.3.1. The tracklet is propagated using the continuous dynamics model associated with the initialized discrete state until it reaches a decision point. We defined three types of decision points: (i) decision to attempt crossing or walk straight ahead without crossing, (ii) decision to accept the currently available gap or not, and (iii) decision on which direction to turn after crossing the street (refer Figure 4.9b). When the current tracklet reaches a decision point, that tracklet branches into two tracklets based on the associated probabilities of the future discrete states. There could be multiple decision points with the prediction horizon, and the prediction branches into two at every decision point resulting in multiple possible trajectories. The number of possible trajectories, $\alpha$, is one more than the number of such decision points. Some of these decision points trigger discrete state transitions: approach to walkaway, approach to wait, wait to cross, etc. In contrast, others, such as the decision on the direction after crossing, result in multiple trajectories for the same "walkaway" discrete state.

**Algorithm 3** Real-time trajectory prediction using IHP model.

Initialize Tracklet $X_0, Q_0, \lambda_0^m, \psi_0^{ip}$
**while** $t_i < T$ **do**
  **procedure** PREDICT
    **for** $k \leftarrow 1$ to $N$ **do** ▷ For entire prediction horizon
      **for** $j \leftarrow 1$ to $N_t$ **do** ▷ For all tracklets
        **if** Reached Decision Point **then**
          Predict decision probability given $\lambda_j^m, \psi_j^{ip}$
          Create new tracklets (branches)
          Predict $Q_k$
        **end if**
        **if** Discrete State Transitioned ($Q_k \neq Q_{k-1}$) **then**
          Reset $X_k$ corresponding to $Q_k$
        **end if**
        Predict $X_{k+1} = f_{Q_k}(X_k, \mathcal{N}(0, W))$
        $j \leftarrow j + 1$
      **end for** ▷ end of all tracklets
      $k \leftarrow k + 1$
    **end for** ▷ end of prediction horizon
  **end procedure**
  **procedure** UPDATE
    $X_{i+1} \leftarrow$ observation $\hat{O}_{i+1}$
    $Q_{i+1} \leftarrow X_{i+1}, T(Q_{i+1}|Q_i)$
  **end procedure**
  $i \leftarrow i + 1$
**end while**

In addition to the tracklets formed by the decisions made by the pedestrians, we considered a tracklet with a non-zero probability that does not have a discrete state but always followed constant velocity dynamics, similar to the baseline con-

stant velocity model. This ensured that the predictions from the IHP model were performance-bounded by the constant velocity predictions, i.e., the model would perform at least as well as the constant velocity model. We use the parameters defined in Table 5.3 for real-time trajectory prediction.

Table 5.3: IHP model parameters.

| **Parameter** | **Value** |
|---|---|
| Heading threshold, $\theta_h$ | $45\,^\circ$ |
| Decision zone length, $D$ | $3\,m$ |
| Interaction region height, $I_h$ | $10\,m$ |
| Interaction region angle, $I_\theta$ | $120\,^\circ$ |
| Prediction horizon, $N$ | $1 - 6\,s$ |
| Sampling time, $\Delta t$ | $0.2\,s$ |
| Maximum pedestrian velocity, $v_{max}$ | $2.5\,m/s$ |
| Prediction envelope probability threshold, $\delta_k$ | 0.01 |

### 5.3.6 Interaction region evaluation

Not all pedestrians in the scene affect the POI. The pedestrian must be within the interaction region of the POI. This interaction region is, however, currently unclear. Existing studies either consider this region based on the observable horizon of the pedestrian (≈50 m) [73] or arbitrarily assume a value (e.g. 3 m) [28, 71, 83]. This value is normally chosen based on the complexity of the prediction algorithms to balance the number of pedestrians and the computation cost.

We aim to identify this interaction region (i.e., interaction height and angle combination) based on the crossing decision model performance. We define several combinations of interaction distance ($h$ = 3, 5, 10, 15, and $20\,m$) and interaction angle ($\theta$ = 60°, 90°, 120°, 150°, and 180°. Refer Figure 5.3a). We process the gap data for each combination, develop the crossing decision model, and evaluate its performance.

We are interested in identifying a limiting interaction region and thus considered instances where there were at least one pedestrian interacting with the POI. Though the number of instances when a gap starts remain the same throughout the dataset irrespective of the interaction region parameters, the number of instances when there are at least one interacting pedestrian depends on the size of the interaction region. If the region is too small, neighboring pedestrians are less likely to fall within the interaction region and if the region is too large, pedestrians that are far away and

probably have no influence on the POI's behavior are likely to be considered to predict the POI's crossing decision. The amount of gap data available thus varied based on the parameter combination as too close and narrow zones would not have many interactions.

### 5.3.7 Baseline models

Similar to Section 4.5.7, we compared the performance of the crossing intent and crossing decision models with standard logistic regression baselines, which is a common approach for modeling pedestrian behaviors [43, 169]. We compared our IHP model against a baseline model with pure constant velocity dynamics without discrete categories or switching dynamics. We also compared the IHP model with the MHP model discussed in Chapter IV and with the state-of-the-art Trajectron++ model [28].

The Trajectron++ model predicts the velocity of pedestrians and assumes that the target velocity follows a Gaussian Mixture Model (GMM). The model uses a neural network framework to predict the GMM parameters using a spatiotemporal graph. The graph's nodes represent the road agents (vehicles and pedestrians), and the edges represent the interactions among the road agents. Road agent history is abstracted using Long Short-Term Memory networks (node LSTM) at the nodes. Road agent interactions use the position information of the agents corresponding to the two nodes in the edge. For each node representing a road agent, the relative positions of other neighboring road agents (connected by the graph edges) are calculated. The model assumes that an aggregated pedestrian agent and an aggregated vehicle agent would represent the equivalent interaction effects of all the neighboring pedestrians and vehicles respectively.

The relative positions of all neighboring agents of the same type (car or pedestrian) are aggregated by summing, and the aggregated states are abstracted using an LSTM network (edge LSTM) for each type of agent (car and pedestrian). The decoder neural network takes as inputs the encoded (abstracted) information from the node and edge LSTMs and vectorized map information to produce the velocity predictions. The decoder has a conditionally variational autoencoder (CVAE) that produces different outputs based on a latent variable. The distribution of the latent variable is learned from data. Velocity predictions are made by first randomly sampling a latent variable from the learned distribution and using the sampled latent variable to predict the target velocity. The predicted target velocity is forward propagated (with single integrator dynamics) to produce dynamically feasible trajec-

tories. The Trajectron++ model predicts multiple trajectories, one for each latent variable sampled. The number of trajectories to be sampled is arbitrarily fixed and mainly depends on the computation power available. [28] demonstrated the Trajectron++ model for twenty samples. To maintain consistency, we also evaluated the Trajectron++ model for the same number of samples.

### 5.3.8 Evaluation metrics

We evaluated the classification performance of the crossing intent and crossing decision models. Predictions were assigned to a class when the probability of the predicted class was greater than 0.5. We considered the standard classification metrics such as accuracy, precision, recall, and F1-score.

From the perspective of the AVs, it is more critical to predict pedestrian behavior before and during crossing than when they are just walking on the sidewalk (without crossing). Thus, we evaluated the performance of the models for pedestrians that crossed the street. The predictions were made for the time steps until they had finished the crossing as pedestrian behaviors after crossing are less relevant to the AVs. The real-time trajectory prediction performance was evaluated for the above situations using the final displacement error (FDE) metric. Additionally, we evaluated the IHP model performance with the probabilistic metrics of expectation of ground truth (EGT) and the forward reachable set (FRS) ratio. A representative illustration of the EGT and FRS ratio metrics is shown in Figure 4.10 in Chapter IV.

***Final displacement error (FDE)***: distance between predicted and actual position at the end of prediction horizon.

***Expectation of ground truth (EGT)***: expectation that the ground truth is covered by the prediction envelope for horizon $N$.

***Forward reachable set (FRS) ratio***: the ratio of the size of the prediction envelope to the full forward reachable set.

Consider the ground truth position at time step $i$ to be given by $(x_i, y_i)$. Similarly, the predicted position corresponding to the future $j$ is given by $(\hat{x}_i^j, \hat{y}_i^j)$, where $j \in [1, \alpha]$ and $\alpha$ is the number of possible futures. FDE is calculated individually for each future. The FDE corresponding to the future $j$ is thereby given as,

$$FDE_j = \|(x_N, y_N) - (\hat{x}_N^j, \hat{y}_N^j)\|^2. \tag{5.3}$$

The environment where the pedestrians are walking (roads, sidewalks, crosswalks, etc.,) is discretized into $0.2 \times 0.2\,m$ grids (approximately the space occupied by a standing pedestrian). The IHP model, at every time step, predicts all possible trajectories where the probabilities of the trajectories are calculated from the probabilities of the intents and decisions. The continuous states are estimated using a Kalman filter with a constant velocity prediction model. For each trajectory, the position at each time step is distributed as Gaussian noise $W$ (refer Equation 4.7). The noise is propagated for the entire prediction horizon $N$ through the Kalman filter. As shown in Figure 4.10, each trajectory had an associated prediction envelope ($e_i^j$) given by the variance (Gaussian noise) propagated by the Kalman filter for that trajectory at time step $i$. All the individual prediction envelopes combined to form the prediction envelope for that time step, $E_i = \{e_i^1, e_i^2, ...., e_i^\alpha\}$. Only the grids with a minimum probability threshold of $\delta_k$, similar to [170], were considered to be part of the prediction envelope.

EGT metric is the average of a Boolean variable, $\nu$, at every prediction time step averaged over the entire prediction horizon. $\nu$ is '1' if the ground truth position at a particular prediction time step is within the prediction envelope of that time step and '0' otherwise. EGT is thereby defined as,

$$EGT = \frac{\sum_{i=1}^{N} \nu_i}{N}, \tag{5.4}$$

$$\text{where} \quad \nu_i = \begin{cases} 1 & \text{if } (x_i, y_i) \in E_i \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

For computing the FRSR metric, the fully forward reachable set, $FRS$, of the pedestrian is first calculated. The maximum velocity of the pedestrian is assumed to be $2.5\,m/s$ (estimated from the inD dataset) and a circular $FRS$ is calculated (refer Figure 4.10). FRSR at the end of the prediction horizon $N$ is given as,

$$FRSR = \frac{E_N}{FRS_N}. \tag{5.6}$$

### 5.3.9 Dataset description

We developed and tested the IHP model on the nuScenes dataset [40]. It is a large-scale dataset for autonomous driving with 850 labeled scenes collected from Boston and Singapore. The dataset was collected from an instrumented vehicle that traveled around these cities. Each scene is annotated at 2 Hz ($\Delta\,t = 0.5$s) and is

20 s long, containing up to 23 semantic object classes and HD semantic maps with 11 annotated layers. We split it into training (700 scenes, $\approx 80\,\%$), and testing (150 scenes, $\approx 20\,\%$) sets. The training set was used to train the crossing decision and crossing intent models. The testing set was used to evaluate the performance of the crossing decision and crossing intent models and the real-time trajectory predictions.

We processed the dataset to compile the features for crossing intent (refer Table 5.1) and crossing decision (refer Table 5.2) models for pedestrians approaching or nearby unsignalized crosswalks. The features ($\lambda_m$) for the crossing intent model had an observation window of $3\,s$ and were compiled every $1\,s$ (rolling window) for the duration when a pedestrian was approaching an unsignalized crosswalk. Each observation window had an associated class—'intend to cross', and 'do not intend to cross'—where a pedestrian is considered to intend to cross if they started crossing at any time (even outside the current observation window) during the time they were observed by the EgoV (irrespective of the actions or even the presence of the InterV). Features and their associated class were calculated from the observations of 5630 pedestrians (1185 pedestrians in the test set) while approaching a crosswalk—1970 (385 in the test set) of them having the intent to cross and 3660 (712 in the test set) without the intent to cross—resulting in the cross intent dataset.

The dataset was also processed to identify traffic gaps, both accepted and rejected gaps, for each of the 850 scenes. Instances of traffic gaps were identified when a pedestrian was inside the decision zone and when a traffic gap had just started. Every such instance set had an associated class—'accept gap', and 'reject gap'—where a pedestrian is considered to accept the gap if they started crossing at any time during the time the were interacting with the InterV corresponding to this traffic gap and reject the gap otherwise. The crossing decision model features ($\psi_{ip}$, refer Table 5.1) and their associated class were compiled for these time instances—434 accepted gaps (366 in the train set and 68 in the test set) and 1384 rejected gaps (1182 in the train set and 202 in the test set)—resulting in the gap dataset.

The trajectory information (ground truth) of the pedestrians in the test set was used to evaluate the real-time trajectory prediction performance of the IHP model.

## 5.4 Results and discussion

In Sections 5.4.1 and 5.4.2, we report the performance of the crossing intent and crossing decision models respectively. In Section 5.4.3, we report the results of the interaction region evaluation. In Section 5.4.4, we report the trajectory prediction

results of the IHP model, and finally in Section 5.4.5, we present the computation performance of the IHP model.

### 5.4.1 Crossing intent

The crossing intent model was evaluated on the nuScenes dataset. We used the crossing intent features ($\lambda^m$) compiled from the nuScenes dataset. After testing the performance of various kernel functions (ranging from the simple kernels to more complicated kernels) for the SVM classifier—linear, quadratic, cubic, Gaussian—we chose the Gaussian Kernel for the SVM model since it had the highest performance in terms of F1-score. The training data were used to develop the model, and the developed model was evaluated on the unseen testing data. For an observation duration of $3\,s$, the model performance is shown in Table 5.4. As shown in Table 5.4, the SVM model performed better, with higher accuracy and F1-score, than a baseline logistic regression model in predicting the crossing intent of approaching pedestrians and was used for pedestrian trajectory prediction.

Table 5.4: Crossing intent model performance for the nuScenes dataset for identifying Cross (C) and Not Cross (NC) intents.

| **Model** | **Accuracy** | **Precision (C/NC)** | **Recall (C/NC)** | **F1-Score (C/NC)** |
|---|---|---|---|---|
| Logistic Regression | 0.71 | 0.78/0.66 | 0.63/0.80 | 0.70/0.72 |
| Support Vector Machine | **0.79** | **0.87/0.72** | **0.70/0.88** | **0.78/0.79** |

### 5.4.2 Crossing decision

We used the SVM crossing decision model for predicting the crossing decision. For model training and testing, we used the features extracted from the nuScenes dataset. We recall that the crossing decision training dataset contained feature sets corresponding to 366 accepted gaps and 1182 rejected gaps. To compensate for this imbalance in the two classes, we performed bootstrapping and randomly sampled another 366 feature sets from the feature sets corresponding to the 366 accepted gaps. We also downsampled the rejected gap data by randomly sampling 732 feature sets (equal to the number of accepted gaps after bootstrapping) from the feature sets corresponding to the 1182 rejected gaps. After testing the performance of various kernel functions (ranging from the simple kernels to more complicated kernels) for the SVM classifier—linear, quadratic, cubic, Gaussian—we chose the Gaussian Kernel for

the SVM model since it had the highest performance in terms of F1-score. The model that was developed using the training data was then evaluated on the unseen testing data. We compared the model with a baseline logistic regression model, similar to [43] and an SVM model without the interaction features. As shown in Table 5.5, the SVM model performed better, with higher accuracy and F1-score than the baseline, and was used for pedestrian trajectory prediction.

Table 5.5: Crossing decision model performance for nuScenes dataset for accepting (C) and not accepting (NC) a gap.

| **Model** | **Accuracy** | **Precision (C/NC)** | **Recall (C/NC)** | **F1-Score (C/NC)** |
|---|---|---|---|---|
| Support Vector Machine (including interaction features) | **0.80** | **0.69/0.82** | **0.71/0.82** | **0.70/0.82** |
| Logistic Regression (including interaction features) | 0.60 | 0.45/0.71 | 0.53/0.64 | 0.49/0.67 |
| Support Vector Machine (w/o interaction features) | **0.68** | **0.46/0.83** | **0.66/0.68** | **0.55/0.75** |

We evaluated if the effect of including the interaction features was statistically significant in the crossing decision prediction performance. We used McNemar's test [177] for comparing model performance through accuracy. A contingency table was constructed (refer Table 5.6) based on the agreement/disagreement between the two models: (i) with the three interaction features (interaction distance, interacting pedestrian state, and the group size), and (ii) without the interaction features.

Table 5.6: Contingency table for McNemar test for model comparison. CD-1 is the crossing decision model trained with the interaction features while CD-2 is the model trained without including the interaction features.

| | **CD-1 (Correct)** | **CD-1 (Incorrect)** |
|---|---|---|
| CD-2 (Correct) | a = 464 | b = 68 |
| CD-2 (Incorrect) | c = 134 | d = 112 |

The test calculates a $\chi^2$ test statistic given by $\frac{(b-c)^2}{(b+c)}$. The calculated value of $\chi^2$ = 21.56 is significant ($p < 0.001$). This shows that the two models are significantly different, and thus, including the interaction effects significantly improves the crossing decision prediction performance.

### 5.4.3 Interaction region evaluation

Several combinations of interaction distance and interaction angle were considered as shown in Figure 5.4. Angles typically within a human's field-of-view (120°) and some angles in the peripheral vision were considered. The maximum interaction distance (20 m) was taken as the approximate distance of two pedestrians waiting

| | | Interaction angle | | | | |
|---|---|---|---|---|---|---|
| | | $\theta = 60°$ | $\theta = 90°$ | $\theta = 120°$ | $\theta = 150°$ | $\theta = 180°$ |
| Interaction distance | *d = 3 m* | N = 200<br>P = 0.85/0.99 | N = 222<br>P = 0.86/0.97 | N = 239<br>P = 0.82/0.98 | N = 296<br>P = 0.66/0.85 | N = 300/363<br>P = 0.68/0.87 |
| | *d = 5 m* | N = 264<br>P = 0.80/0.89 | N = 300/307<br>P = 0.75/0.87 | N = 300/337<br>P = 0.75/0.88 | N = 300/386<br>P = 0.76/0.91 | N = 300/445<br>P = 0.71/0.90 |
| | *d = 10 m* | N = 300/463<br>P = 0.78/0.92 | N = 300/507<br>P = 0.72/0.88 | N = 300/536<br>P = 0.74/0.91 | N = 300/570<br>P = 0.68/0.90 | N = 300/612<br>P = 0.69/0.90 |
| | *d = 15 m* | N = 300/586<br>P = 0.68/0.88 | N = 300/625<br>P = 0.70/0.91 | N = 300/654<br>P = 0.64/0.87 | N = 300/686<br>P = 0.65/0.87 | N = 300/723<br>P = 0.58/0.88 |
| | *d = 20 m* | N = 300/633<br>P = 0.67/0.87 | N = 300/683<br>P = 0.69/0.91 | N = 300/716<br>P = 0.65/0.90 | N = 300/748<br>P = 0.63/0.87 | N = 300/782<br>P = 0.62/0.89 |

Model performance in 1st quartile (0.58-0.65)
Model performance in 2nd quartile (0.66 – 0.68)
Model performance in 3rd quartile (0.69 – 0.74)
Model performance in 4th quartile (0.75 – 0.86)

Figure 5.4: The crossing decision data was compiled and an SVM model was trained for each combination. To avoid over-fitting, the training data had a maximum limit of 300 samples (denoted by N). P is the F1-score performance of predicting cross intent/no cross intent. It can be seen that crossing decision prediction performance P was best for interaction distances of 10 m and less and interaction angles of 120° and less.

on the opposite sides of a crosswalk of a four-lane street. Distances below 3 m did not yield sufficient gap data to get meaningful results and were thus neglected. The gap data were processed for each interaction distance and angle combination, and a two-class SVM model (with a Gaussian kernel) was trained. To avoid over-fitting, the overall data was limited to 300 with a training and testing split of 80 % and 20 %.

The different combinations were evaluated using the F1-score of the model evaluated on the 20 % testing data. It can be seen that crossing decision prediction performance (P = F1-score of cross intent/F1-score of no cross intent) was best for interaction distances of 10 m and less and interaction angles of 120° and less. This indicates that the effects of nearby pedestrians are less relevant beyond an interaction distance of 10 m and an interaction angle of 120°.

### 5.4.4 Trajectory prediction

In this section, we discuss the evaluation of the IHP model for predicting pedestrian trajectories using the steps discussed in Algorithm 3. The IHP model was evaluated on the nuScenes dataset. We evaluated the IHP model only on pedestrians close to the crosswalk or who were crossing the street. We did not evaluate the predictions after a pedestrian has crossed a street as this is less relevant from the AV's perspective.

For deterministic error metrics, we compared the best prediction trajectory (in terms of least FDE) with the ground truth trajectory. The best prediction of the IHP model would be the same as the best prediction from the MHP model. This is because both these models have the same decision points. The difference arises in the probability of the decisions, and therefore, we expect differences between these models in the probabilistic metrics.

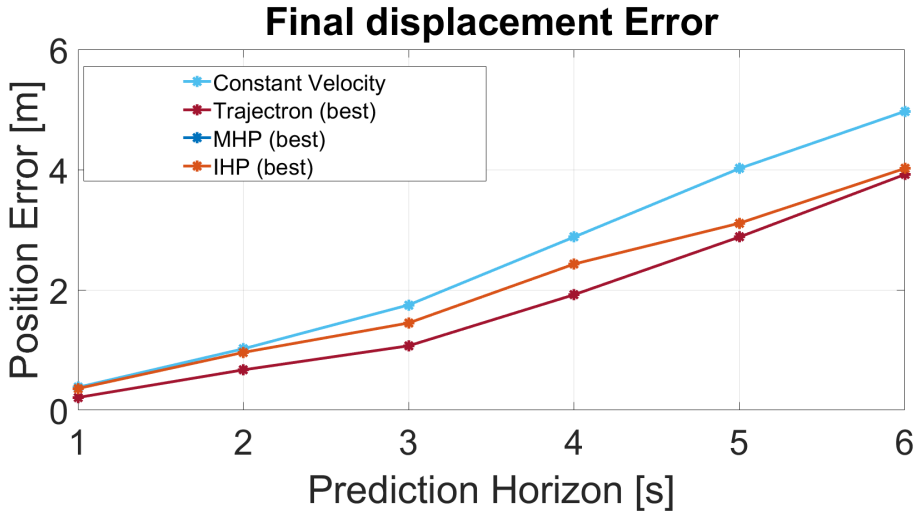
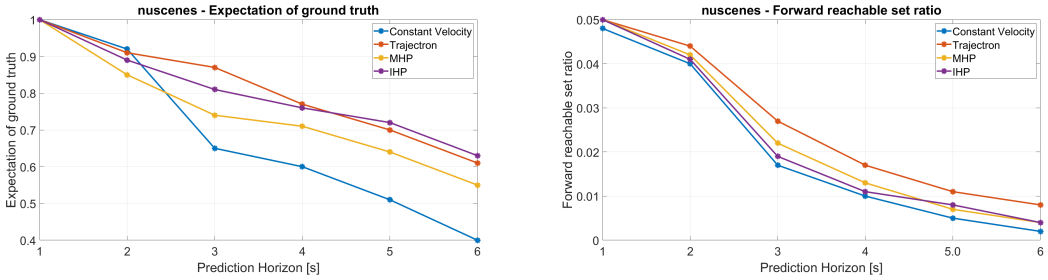



Figure 5.5: Pedestrian trajectory prediction comparison in nuScenes dataset at different prediction horizons for (a) average displacement error metric, and (b) final displacement error metric. The best prediction from the Trajectron++ model is better than the best prediction from the IHP model. The performance difference is however, not large.

From Figure 5.5, it can be observed that though the best prediction of the IHP model (and the MHP model) performed better than the baseline constant velocity

model baseline but did not perform as well as the best prediction from the Trajectron++ model.




(a) (b)

Figure 5.6: Comparison of (a) expectation of ground truth and (b) FRS ratio of the IHP model with the baseline model and Trajectron++ model. The ground truth trajectory is more likely to be captured in the Trajectron++ model at smaller time steps. With increased prediction horizon, the IHP model better captures the ground truth. The relatively higher value of FRS ratio in the IHP model indicates a smaller prediction envelope than the Trajectron++ and baseline models.

We evaluated the performance of the IHP model on the nuScenes dataset through probabilistic metrics. Figure 5.6 shows the expectation of ground truth and FRS ratio for various prediction horizons. The expectation of ground truth gives the likelihood that the ground truth trajectory is in the set of multimodal predicted envelopes. It can be observed that the probability was higher at lower time steps and decreased at higher time steps. The Trajectron++ model was more likely to capture the ground truth at smaller prediction horizons, whereas the IHP model is more likely to capture the ground truth at longer horizons.

The IHP model was more likely to capture the ground truth than the baselines for all time horizons. The IHP model also had a lower FRS ratio. A high FRS ratio indicates more conservative behavior because the prediction envelopes include a higher proportion of possible behaviors. The Trajectron++ model has a higher FRS ratio because it predicts up to 20 futures, whereas the maximum number of futures predicted in the IHP model is 10. Thus IHP model better captured the ground truth without being overly conservative.

The IHP model efficiently predicted multimodal trajectories for the long term. Unlike the state-of-the-art methods [27, 28, 97], the multimodality was not arbitrarily conditioned on a latent variable but instead was conditioned on the decision-making

process of the pedestrian, which is grounded in actual pedestrian behavior and thereby easier to comprehend.

### 5.4.5 Computation performance

The experiments were run on Intel Core i7-7700K, 16 GB RAM, and Nvidia GTX 1080 Ti. On average, the IHP algorithm took 0.09 $s$ to run predictions for the entire prediction horizon of up to 6 $s$, whereas the Trajectron++ model took 0.21 $s$ for the same prediction horizon. This difference in performance is primarily due to the intermittent inference of the IHP model compared to the continuous inference of the Trajectron++ model. The crossing intent and crossing decision SVM models were inferred only when the pedestrians reached the corresponding decision points. During the rest of the time, the IHP model ran simple constant velocity models. This enabled the model to run for long-term and multiple predictions independently for each pedestrian in the scene.

## 5.5 Chapter Summary

The contribution presented in this chapter is the extension of the MHP model discussed in Chapter IV to incorporate the interactions with nearby pedestrians. The developed IHP model had a similar multimodal performance as the state-of-the-art Trajectron++ model for situations relevant to the AVs, such as when approaching the crosswalk to cross or crossing the street. The IHP model, on the other hand, was more intuitive and computationally efficient than the Trajectron++ model. The model was validated in the nuScenes dataset.

# CHAPTER VI

# Behavior-aware Automated Vehicle Control

## 6.1 Overview of chapter

The contribution discussed in this chapter is the development of an AV controller for safe and trustworthy motion planning that incorporates the predictions from the pedestrian behavior models. The controller demonstrates the ability to design driving behaviors that the pedestrians can potentially understand, which could facilitate safe and trustworthy navigation. This chapter discusses the development and validation of a behavior-aware controller based on model predictive control (MPC) framework, which is also published in [30, 41].

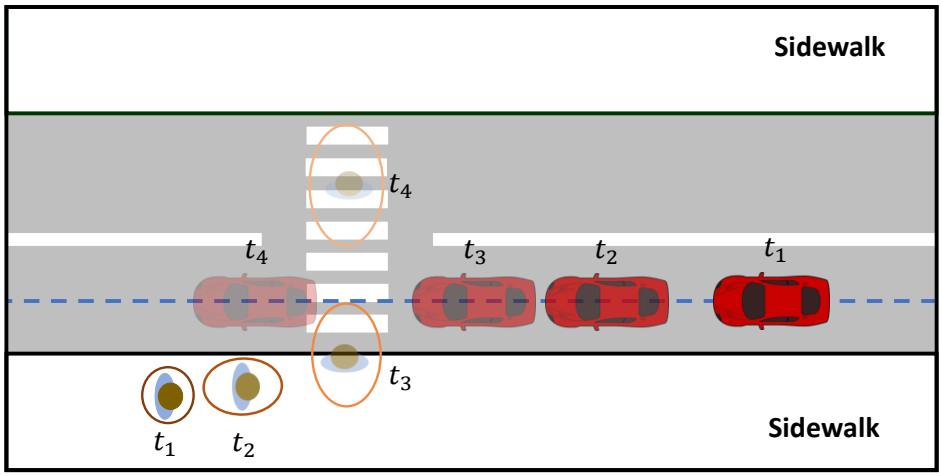



Figure 6.1: Representation of a typical interaction between a vehicle and a crossing pedestrian. The vehicle has to plan its trajectory considering the moving pedestrian and by following the road's centerline. The illustration shows the predicted pedestrian trajectory and the uncertainty ellipses at various time instances and the vehicle's planned trajectory.

## 6.2 Behavior-aware model predictive control

Figure 6.1 illustrates the AV behavior-aware control problem for a typical interaction with a pedestrian at an unsignalized mid-block crosswalk. Studies have incorporated pedestrian behavior into AV control by assuming pedestrians as moving obstacles with a constant velocity and Gaussian noise [178, 179]. This simple approach is effective for short duration planning ($\approx$ 1-2 $s$) and when the pedestrians are moving. However, pedestrian behavior is much more unpredictable at crosswalks as they have to wait for an opportunity and decide when to cross.

Consider the scenario where an AV is approaching an unsignalized crosswalk as shown in Figure 6.1. Pedestrians approaching the crosswalk decide to either cross the road or wait for the AV to pass. The AV has to plan actions that ensure safety and help the riders in the AV reach their destinations comfortably. Using the pedestrian crossing model discussed Section 4.3 and in [35], AVs can predict future pedestrian states and plan their actions accordingly. In the following, we explain the AV and pedestrian models used and formulate a receding horizon model predictive control problem to calculate the AV control inputs.

### 6.2.1 AV model

Dynamic vehicle models are comprehensive but challenging to use for real-time AV motion planning, especially in urban scenarios. In addition to being computationally expensive, the tire models have vehicle velocity in the denominator for tire slip estimation and become singular at low speeds. Hence, these models are not suitable for stop-and-go scenarios common in urban driving [180]. Thus, we assume the AV to be a point mass with a rectangular footprint (refer Figure 6.2). We assume that longitudinal vehicle dynamics are sufficient for the crosswalk interactions and employed a discrete-time kinematic model shown in equation (1), where $\mathbf{X} = [x_v \; v_v]^\top$ is the state vector comprising the position and velocity of the vehicle respectively. $\Delta\, t$ denotes the discretization time step, and $a_v$ is the acceleration input that governs the AV's motion.

$$\begin{aligned} x_{v_{k+1}} &= x_{v_k} + \Delta t \, v_{v_k} \\ v_{v_{k+1}} &= v_{v_k} + \Delta t \, a_{v_k} \end{aligned} \tag{6.1}$$

### 6.2.2 Problem formulation

The behavior-aware model predictive controller (B-MPC) calculates the inputs to achieve the AVs' objectives expressed through a cost function. The physical limitations of the AV and collision avoidance with pedestrians are expressed as constraints. The problem is formulated as a constrained quadratic optimization problem, which enables fast computation of control inputs, suitable for real-time planning. The optimization problem is formalized in equation (2), where $J$ is the cost function and $\mathbf{Z} = [\mathbf{X}\ \mathbf{V}\ \mathbf{U}\ \Delta\mathbf{U}\ \mathbf{R}]^\top$ is a stacked vector of all states, control inputs, change in control, and references, respectively, for horizon $N$.

$$\begin{aligned} \min_{\mathbf{Z}} \quad & J(\mathbf{Z}) \\ \text{s.t.} \quad & A_{eq}\,\mathbf{Z} = B_{eq} \\ & A_{ineq}\,\mathbf{Z} \leq B_{ineq} \\ & l_b \leq \mathbf{Z} \leq u_b \end{aligned} \tag{6.2}$$

### 6.2.3 Cost function

Safety is the main priority in the AV control problem. However, the AVs should also follow speed limits and reach their destination on time while maintaining ride comfort. The quadratic cost matrices $Q$ for the various objectives are constructed using their corresponding weights $w$, chosen to be positive to ensure the matrices $Q$ are positive semi-definite. The objective cost function is given by

$$J = J_{target} + J_{jerk} + J_{acc} + J_{speed}. \tag{6.3}$$

#### 6.2.3.1 Target cost

One of the primary objectives of the AV is to reach a target destination. Since pedestrians always cross the street at the crosswalk (refer Assumption IV.5), for the purpose of simplicity, we consider the destination to be an arbitrary point $x_v^{ref}$ beyond the crosswalk (refer Fig. 6.4). This ensures that passing the crosswalk is one of the objectives of the AV. The difference between the destination $x_v^{ref}$ and the vehicle position at the end of the prediction horizon $x_v^N$ is penalized as $J_{target} = (x_v^N - x_v^{ref})\, Q_{target}\, (x_v^N - x_v^{ref})$, where, $Q_{target} = w_{target}$.

#### 6.2.3.2 Comfort cost

The other objective of the AVs is to ensure ride comfort for the people inside the vehicle. Ride comfort is typically characterized by the jerk of the vehicle. Both sudden acceleration and sudden deceleration reduce ride comfort. Thus we penalize sudden changes in acceleration as $J_{jerk} = \Delta\,\mathbf{U}^{\top}\,\mathbf{Q_{jerk}}\,\Delta\,\mathbf{U}$. Moreover, the acceleration is also penalized to restrict unnecessary acceleration or deceleration by the vehicle as $J_{acc} = \mathbf{U}^{\top}\,\mathbf{Q_{acc}}\,\mathbf{U}$. The above quadratic costs are given by $\mathbf{Q_{jerk}} = diag(w_{jerk}, ...., w_{jerk})$, and $\mathbf{Q_{acc}} = diag(w_{acc}, ...., w_{acc})$.

#### 6.2.3.3 Speed cost

AVs are expected to follow the posted speed limit to maintain a smooth flow of traffic. Thus, we penalize the deviation from the reference speed (both higher and lower speeds) as $J_{speed} = (\mathbf{V} - \mathbf{V_{ref}})^{\top}\,\mathbf{Q_{speed}}\,(\mathbf{V} - \mathbf{V_{ref}})$, where, $\mathbf{Q_{speed}} = diag(w_{speed}, ...., w_{speed})$.

### 6.2.4 Constraints

AV motion is constrained to follow the model discussed in equation 6.2.1. States and inputs are also constrained considering the physical limitations of the vehicle and to avoid potential collisions with pedestrians. The different constraints developed are discussed below.

#### 6.2.4.1 AV motion model

To ensure that the optimization problem calculates states and inputs that physically agree with the motion of the vehicle, the motion model mentioned in equation 6.2.1 is given as equality constraints.

#### 6.2.4.2 State and control bounds

Considering the physical limitations of the vehicle, we restrict the velocity, acceleration, and jerk of the vehicle represented as lower ($l_b$) and upper bounds ($u_b$) in equation 6.2.2. The speed bound is higher than the speed limit. This enables the AVs to choose higher speeds, if needed, to satisfy the constraints.

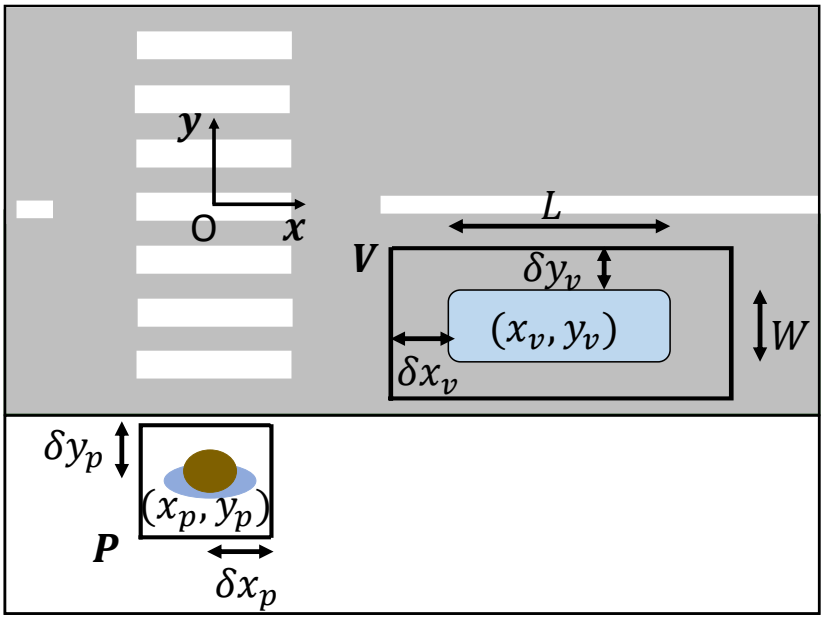



Figure 6.2: Collision avoidance is incorporated by ensuring the sets **P** and **V**, representing the uncertain positions of AV and pedestrian respectively, do not intersect each other for the entire planning horizon.

#### 6.2.4.3 Collision avoidance

Pedestrian trajectory is predicted using the pedestrian crossing model discussed in Section III-C. Collision avoidance of the planned AV trajectory with the predicted pedestrian trajectory is ensured through inequality constraints in the optimization problem (refer to equation 6.2.2). We incorporate the uncertainty in the state estimation of pedestrians and vehicles as over-approximated rectangles, which is a conservative assumption that ensures safety. To avoid collisions, the sets **P** and **V** (refer Fig. 6.2) should not intersect with each other at any time instant. This is expressed by the sets of inequality constraints in both $x$ and $y$ axes, represented by equations 6.2.4.3 and 6.2.4.3 respectively. A collision is avoided if at least one of the following four equations is satisfied at any given time.

$$\begin{aligned} x_p + \delta x_p &\leq x_v - L/2 - \delta x_v \\ x_p - \delta x_p &\geq x_v + L/2 + \delta x_v \end{aligned} \tag{6.4}$$

$$\begin{aligned} y_p + \delta y_p &\leq y_v - W/2 - \delta y_v \\ y_p - \delta y_p &\geq y_v + W/2 - \delta x_v \end{aligned} \tag{6.5}$$

We assume the AVs can accurately track the centerline of the lane and thus neglect the lane boundary conditions in our formulation. The B-MPC can effectively combine pedestrian crossing behavior predictions for long durations ($> 5\,s$) as constraints for collision avoidance and calculate inputs that optimize the AV's objectives.

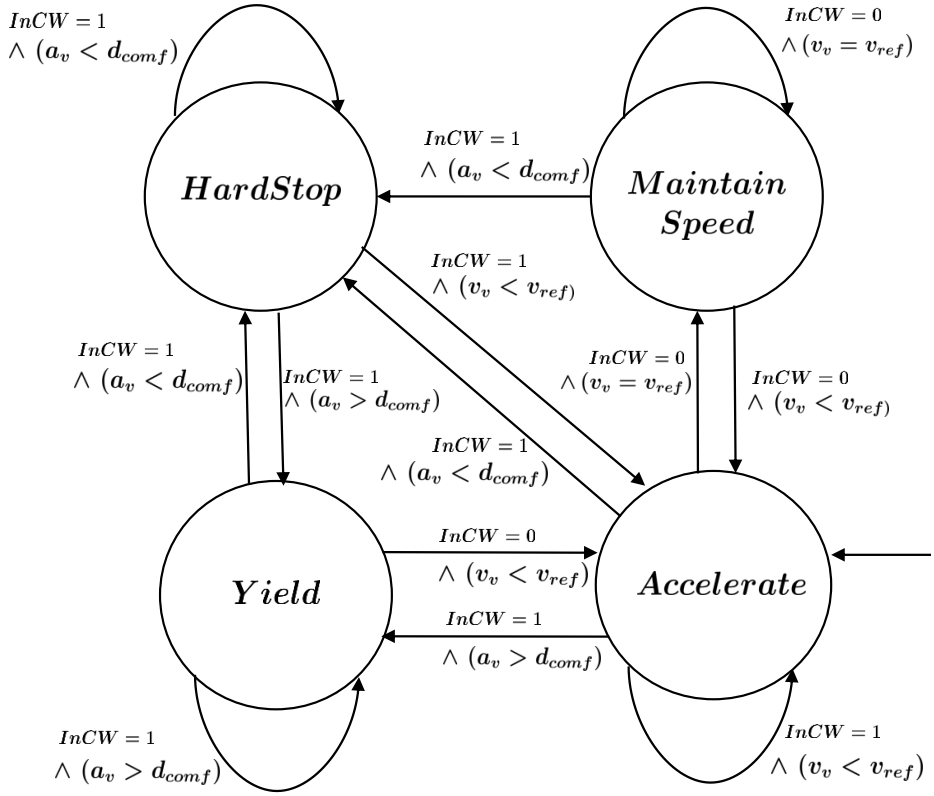



Figure 6.3: Baseline Rule-based controller.

## 6.3 Baseline controller

The developed B-MPC is compared against a baseline rule-based controller. The baseline controller is a simple finite state machine (FSM) with four states: *Maintain Speed*, *Accelerate*, *Yield*, and *Hard Stop* (refer Fig. 6.3). The Boolean variable *InCW* denotes the pedestrian's crossing activity. *InCW* is 'one' from when the pedestrian started moving laterally to cross until they completely crossed the AV lane, and 'zero' otherwise. The variable $d_{comf}$ denotes the comfortable deceleration limit. The controller normally maintains the speed limit, $v_{ref}$. Whenever a pedestrian starts walking to cross the road, the controller always tries to stop, either by yielding or through a hard stop. The deceleration is calculated as $a_v = -\frac{{v_v}^2}{2\,dist_{stop}}$, where $v_v$ is

the vehicle's current velocity and $dist_{stop}$ is the distance available to the AV before which it has to stop to avoid a collision. The stopped vehicle then accelerates back to its nominal speed once the pedestrian has crossed the AV's lane. The increments in the acceleration and deceleration at every time step are controlled by the comfortable jerk limits in the *Yield* state and the hard jerk bounds in the *Hard Stop* state. Simulations are run with the same vehicle and pedestrian parameters shown in Table 6.1.

Table 6.1: Parameters used in the simulation.

| **Parameter** | **Value Range** |
|---|---|
| Vehicle spawn speed $[m/s]$ | 14 to 16 |
| Speed limit, $v_{ref}$ $[m/s]$ | 16 |
| Spawn time gap between vehicles, $t_{spawn}$ $[s]$ | 1 to 8 |
| Minimum time gap between vehicles to avoid collision, $t_{min}$ $[s]$ | 2 |
| Hard speed bounds $[m/s]$ | 0 to 50 |
| Comfortable acceleration limits, $d_{comf}$, $a_{comf}$ $[m/s^2]$ | -5 to 2 |
| Hard acceleration bounds $[m/s^2]$ | -10 to 10 |
| Comfortable jerk limits $[m/s^3]$ | -5 to 2 |
| Hard jerk bounds $[m/s^3]$ | -10 to 10 |
| Pedestrian decision zone length, $d_y$ $[m]$ | -3 to 1 |
| Pedestrian speed $[m/s]$ | 1 to 1.5 |
| AV destination, $P_{ref}$ $[m, m]$ | (-120, -1.75) |
| Prediction horizon, $N$ $[s]$ | 5 |
| Normalized cost function weights, | |
| $w_{target}$ | 0.004 |
| $w_{jerk}$ | 8 |
| $w_{acceleration}$ | 0.02 |
| $w_{speed}$ | 0.01 |

## 6.4 Simulation

To evaluate the performance of the controllers, we simulated a scenario where AVs are approaching an unsignalized mid-block crosswalk with a pedestrian possibly intending to cross the street (refer Figure 6.4).

We simulate a midblock scenario with straight roads and assume the AVs follow the centerline of the lane. We developed the simulation to be as realistic as possible by considering a stream of AVs approaching the crosswalk one after the other with varying speeds and time gaps. However, at any time, only one pedestrian will be in

the simulation. AVs spawn with a random initial speed and a randomly varying time gap between their spawns. This ensures that there are both crossable and uncrossable gaps for the pedestrians. The simulation parameters are shown in Table 6.1. The decision zone $D$ is larger on the side of the approaching pedestrians (refer Table 6.1). This ensures that approaching pedestrians have the opportunity to evaluate a new gap and decide to cross or wait whenever they are within $D$.

AVs assume that all pedestrians approaching the crosswalk have the intention of crossing the street until they walk past the crosswalk and out of the decision zone, $D$ (refer Figure 6.4). However, only a fraction of pedestrians ($\approx 80\%$) is randomly assigned the intention to cross the street. Pedestrians who intend to cross evaluate the gap within the decision zone, whereas others walk past the crosswalk at a constant velocity. Figure 6.4 illustrates the AV – pedestrian interactions in the simulation. The gap of $AV_2$ will start when $AV_1$ has crossed the pedestrian, at which point the pedestrian can decide to cross or wait.

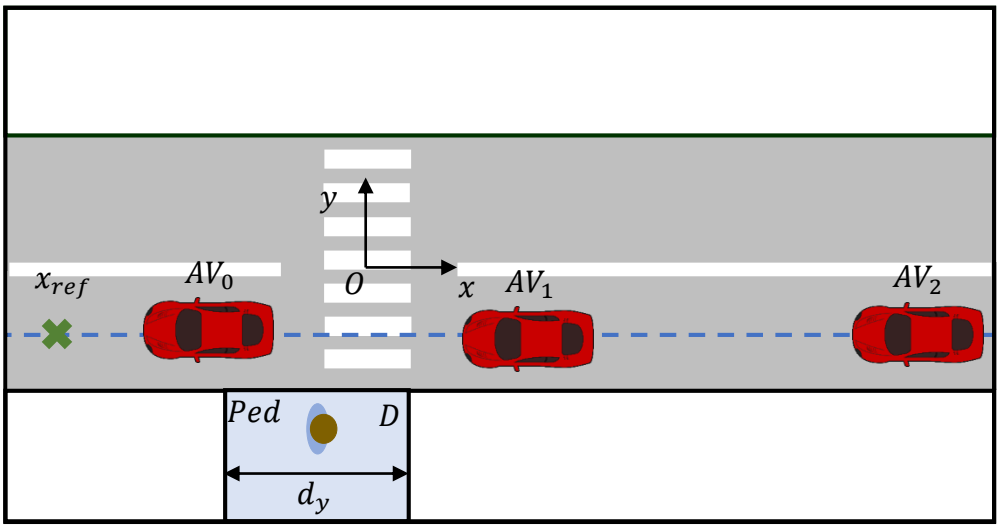



Figure 6.4: Illustration of the AVs interacting with a pedestrian in the simulation. AVs' objective is to reach $x_{ref}$, given in Table 6.1. Decision zone of pedestrians is represented by the set $D$ with length $d_y$.

## 6.5 B-MPC results and discussion

The constrained quadratic control problem was solved using the standard quadratic program solver in MATLAB. The simulation was run for 500 pedestrians for both the B-MPC and the baseline controllers as shown in Table 6.2. The average run time of

B-MPC with the prediction model was 24.7 $ms$ with a standard deviation of 3.1 $ms$.

Table 6.2: Simulation runs for B-MPC and Baseline controllers.

| **Parameter** | **B-MPC** | **Baseline** |
|---|---|---|
| No. of Pedestrians | 500 | 500 |
| No. of AVs | 1434 | 1401 |
| No. of Crossings (Accepted gaps) | 411 | 405 |
| No. of pedestrians without crossing intent | 89 | 95 |

### 6.5.1 B-MPC performance

Fig. 6.5 shows the B-MPC performance at various time instances for a nominal pedestrian interaction. The pedestrian in this case accepted a gap of 3.6 $s$. Initially, the AV travels at its preferred speed. The AV predicts the pedestrian will cross and reacts by starting to slow down at $t = 7.9$ $s$. The AV starts to accelerate at $t = 11.0$ $s$ even before the pedestrian has crossed its lane. Finally, the AV goes past the pedestrian at $t = 13.3$ $s$. The speed changes can be seen through the changes in the spacing of the AV trajectory points (red points in Figure 6.5). The long horizon prediction helps the AV react early to the crossing pedestrian, much before the pedestrian starts walking or crossing the AV lane.

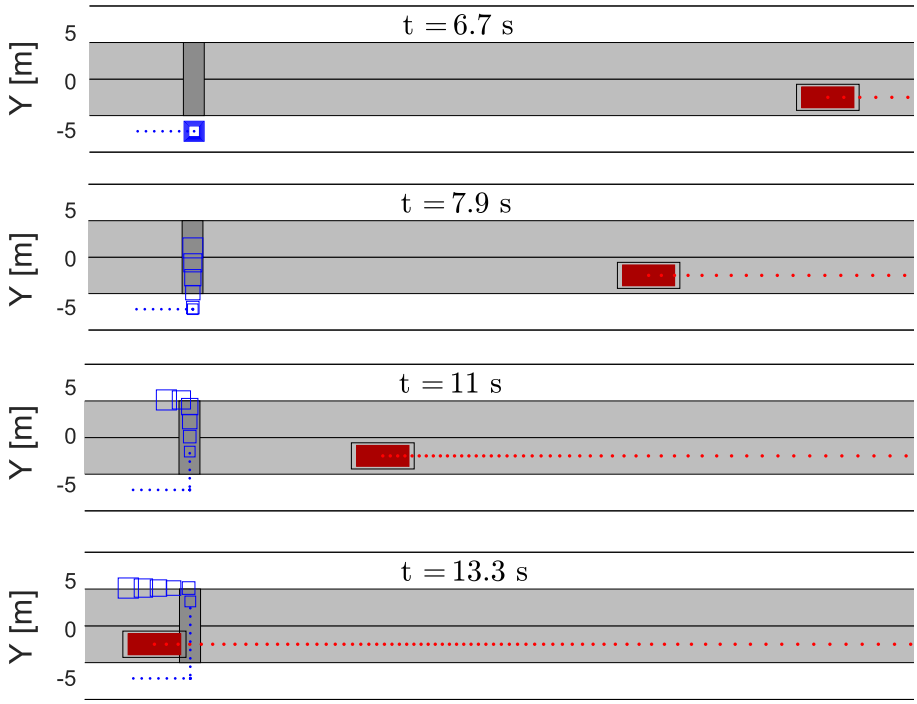



Figure 6.5: A typical interaction between the AV and a pedestrian. The AV is represented by the red rectangle with the black rectangle indicating the position uncertainty. The red and the blue dots indicate the trajectories taken by the AV and the pedestrian respectively and the blue rectangles indicate the predictions of pedestrian trajectory by the AV. The AV starts slowing down at $t = 7.9\ s$ and starts accelerating at $t = 11\ s$, even before the pedestrian crosses the lane.

### 6.5.2 Baseline comparison

We compared the B-MPC and the baseline controllers for varying time gaps and varying AV spawn speeds (refer Table 6.1) for the cases when the pedestrians had the intent to cross. Fig. 6.6 compares the collision avoidance performance between the two controllers by evaluating the minimum distance to pedestrians ($d_{ped}$). It can be seen that the B-MPC has an overall higher minimum distance to the pedestrian than the baseline case. For short gaps, the baseline controller sometimes cannot avoid collisions (4 out of 500 cases). Whereas the B-MPC avoids collisions for the range of gaps simulated. The B-MPC can thus handle a wider range of gaps and is applicable for a wider range of scenarios than the rule-based controller.

Figure 6.7 compares other performance measures between the two controllers such as time to destination ($t_{des}$), average velocity ($v_m$), average acceleration ($a_m$), and average absolute jerk ($j_m$) during the interaction duration. Interaction duration was calculated as the duration between two time instances. The first instance was when

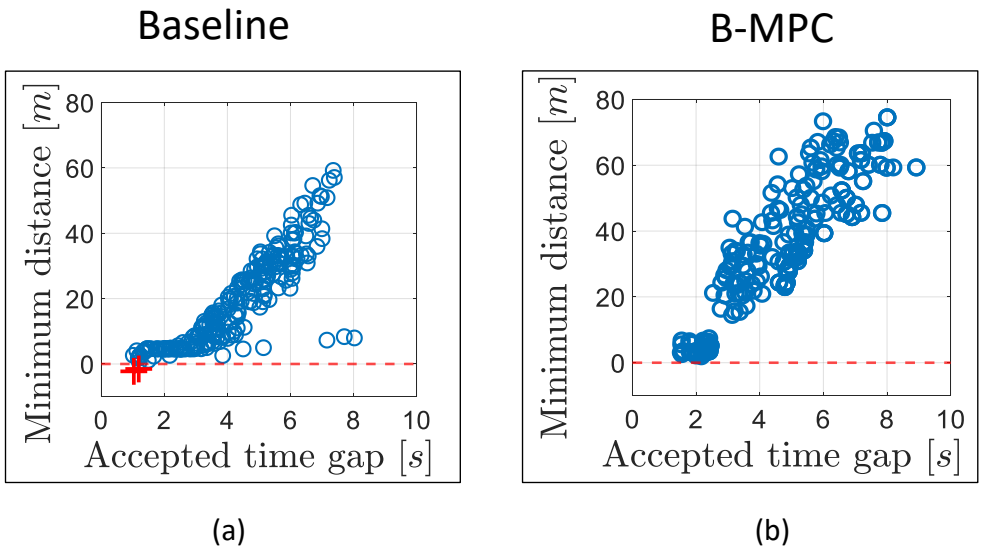



Figure 6.6: Comparison of minimum distance to pedestrians between baseline controller and B-MPC. The red 'plus' marks indicate the instances of collisions between the AV and the pedestrian. B-MPC is able to avoid collisions comfortably whereas collisions are inevitable at shorter gaps for the baseline controller.

the pedestrian started walking to cross, in the case of the baseline controller, or when the AV had predicted the start of pedestrian walking, in the case of the B-MPC. The second instance was when the pedestrian crossed and left the lane in which the AV traveled or when the AV had crossed the pedestrian, whichever occurred earlier. The overall average duration of interaction was higher for the B-MPC: $t = 6.08\ s$ for B-MPC and $t = 4.05\ s$ for the baseline controller. The B-MPC can reach its destination faster as it does not come to a complete stop unless necessary to avoid collisions, thereby improving the traffic flow. It can be seen that the B-MPC is more aggressive, efficient, and comfortable than the baseline, as observed through the higher average velocity, lower average acceleration effort, and lower average jerk, respectively.

### 6.5.3 Non-crossing pedestrian interaction performance

We report the performance of both the B-MPC and the baseline controllers for the cases where the pedestrians did not intend to cross (refer Table 6.3). The B-MPC reacts to the approaching pedestrians within $D$, seen from the slightly higher deceleration and reduced velocity. The baseline controller does not react at all since it never sees the pedestrian crossing laterally. Even still, the overall performance of the B-MPC is better, as seen by the lower deceleration, higher distance to pedestrian, and lower time to destination, than the baseline for our sample case where approximately

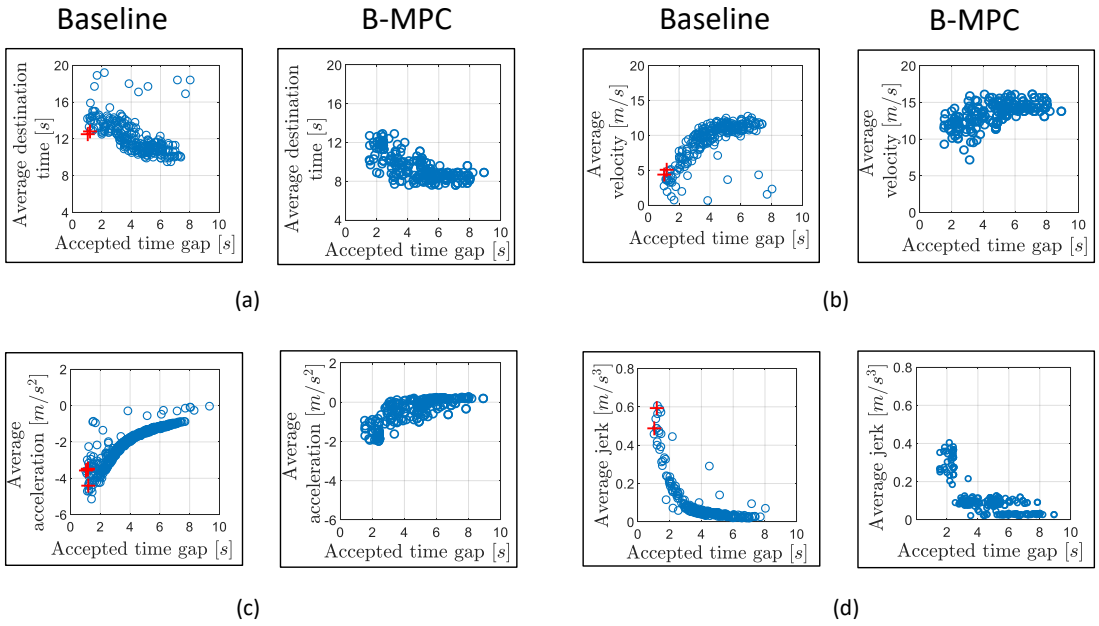



Figure 6.7: Performance metrics comparison between baseline controller and B-MPC. The performance metrics compared are (a) time to destination, (b) average velocity, (c) average acceleration, and (d) average jerk during the interaction duration. The B-MPC is more efficient and comfortable as it results in less time to reach destination, less control effort, and less jerk than the baseline. The red 'plus' signs indicate the instances of collision.

20% of pedestrians approaching the crosswalk did not intend to cross.

Table 6.3: Performance metrics for B-MPC and baseline controllers for pedestrians without crossing intent and all pedestrians.

| **Parameter** | **B-MPC, No intent** | **Baseline, No intent** | **B-MPC, Overall** | **Baseline, Overall** |
|---|---|---|---|---|
| $t_{int}$ $[s]$ | 6.44 | 3.38 | 6.08 | 4.01 |
| $t_{des}$ $[s]$ | 8.74 | 8.27 | 9.47 | 11.14 |
| $d_{ped}$ $[m]$ | NA | NA | 33.7 | 20.1 |
| $v_m$ $[m/s]$ | 13.9 | 14.8 | 13.3 | 9.3 |
| $a_m$ $[m/s^2]$ | -0.10 | -0.04 | -0.51 | -1.90 |
| $j_m$ $[m/s^3]$ | 0.04 | 0.04 | 0.11 | 0.18 |

## 6.6 AV behavior design

AV interactions with pedestrians, especially at uncontrolled mid-block crosswalks, are challenging as the right-of-way in the US varies by state [32]. To ensure safety, AVs are expected to drive cautiously around pedestrians. However, overly cautious AV behavior can encourage pedestrians to be careless or abuse the AVs [61] by not letting the AVs pass. In such cases, the AVs should safely nudge forward until they can pass the pedestrians. Additionally, riders inside the AV expect a comfortable ride and to reach their destinations (AV performance) without excessive delay [181].

These objectives of safety, performance, and comfort are opposing in nature and thus have to be balanced in a way accepted by both the riders and the pedestrians. The existing approach of fixing the weights of the objectives [182, 183] can be constrictive as in reality, the objectives can vary depending on the situation. For example, in the case of a pedestrian crossing at an unsignalized crosswalk, the AV might want to prioritize comfort and gradually slow down to stop. However, at a school zone, the AV might want to prioritize safety and have a stronger deceleration. Further, existing approaches to AV decision-making consider pedestrians as independent moving obstacles [183, 184] and do not explicitly consider the effects of the vehicle behavior on the pedestrian behavior. Pedestrians, in general, perceive a defensive driving style (which prioritizes safety) as more acceptable and are more willing to cross under such situations [34, 64]. However, the objective evaluation of these interactions – gap accepted, average acceleration, average jerk (an indicator of sudden acceleration/deceleration), etc., – of the different driving styles is not known.

We objectively evaluated a spectrum of driving behaviors represented by the different weight combinations of the three AV driving objectives – safety, performance, and comfort. We incorporated the UHP pedestrian model to predict pedestrian trajectories. Our findings suggest the possibility of characterizing driving behaviors by varying the weights of the driving objectives. We believe that by understanding the objective performance of the different driving behaviors with an interactive pedestrian model, the AV-decision making can be expressed by varying the weights of the objective function.

### 6.6.1 Methodology

We used the same simulated scenario as in Section 6.5.3 with straight roads, similar to actual midblocks, and the AVs followed the centerline of the lane as shown in Figure 6.4. We assumed that AVs use only longitudinal control (acceleration/deceleration)

while interacting with crossing pedestrians. Pedestrian crossing behavior was simulated using the UHP model discussed in Section 4.3.

We developed an MPC controller for the AVs that used predictions of pedestrian behavior. We assumed the AV to be a point mass with a rectangular footprint. We employed a discrete longitudinal kinematic model shown in equation 6.2.1, where $x, v, a$ are the longitudinal position, speed, and acceleration inputs of the vehicle, respectively. $\Delta t$ denotes the discretization time step.

### 6.6.2 Cost Function and Constraints

Safety is a main priority in the AV planning and control problem. However, the AVs should also adhere to the speed limits and reach their destination quickly while maintaining ride comfort for the people inside. The optimization objective is formulated as the minimization of a cost function. This cost function serves as an inverse description of the ideal vehicle behavior and consists of three costs – safety, performance, and comfort. The safety cost is a function of the distance of the AV from the pedestrian when the pedestrian is in the same lane as the AV. The performance cost is given by the distance to the target location and the deviation of the AV speed from the speed limit. The comfort cost is given by the magnitude of acceleration and jerk. The cost function is thus represented by,

$$\begin{aligned} J &= J_{safety} + J_{performance} + J_{comfort} \\ J_{safety} &= w_s \, f_s(x_{Ped}, x) \\ J_{performance} &= w_p \, (f_t(x, x_{ref}) + f_v(v, v_{ref})) \\ J_{comfort} &= w_c \, f_c(a, \Delta a) \end{aligned} \tag{6.6}$$

AV motion is constrained to follow the model discussed in equation 6.2.1. States and inputs are also constrained considering the physical limitations of the vehicle. Although the inclusion of the safety cost could result in safe vehicle maneuvers, minimal safety of collision avoidance is enforced by incorporating collision avoidance constraints in addition to the safety cost, similar to [185]. Pedestrian trajectory was predicted using the UHP model for a horizon of 3 seconds.

### 6.6.3 Implementation

We developed the simulation to be as realistic as possible by considering a stream of AVs approaching the crosswalk one after the other with varying speeds and time gaps. However, at any time, only one pedestrian was in the simulation. We assumed

that all pedestrians had the intention to cross the street. AVs spawned with a random initial speed (13 - 16 m/s) and a randomly varying time gap (1 - 7 s) between their spawns. The different cost functions have scaling factors to ensure similar ranges of the absolute values for $f_s$, $f_c$, and $f_t + f_v$. This ensures that the safety, performance, and comfort costs are similar under the same weights. We evaluated seven combinations of the weights as shown in Table 6.4. Simulations were run 200 times for each of the seven weight combinations. The performance weight was ensured to be non-zero so that the vehicle moves forward, especially after stopping for a crossing pedestrian.

Table 6.4: Different combinations of the three weight parameters.

| **Weight combination** | $w_s$ | $w_p$ | $w_c$ |
|---|---|---|---|
| S | 0.950 | 0.050 | 0.000 |
| P | 0.000 | 1.000 | 0.000 |
| C | 0.000 | 0.050 | 0.950 |
| PC | 0.000 | 0.500 | 0.500 |
| SC | 0.475 | 0.050 | 0.475 |
| SP | 0.500 | 0.500 | 0.000 |
| SPC | 0.333 | 0.333 | 0.333 |

### 6.6.4 Behavior design results and discussion

The objective evaluation of the various controllers with the different weight combinations is shown in Table 6.5. Collisions were observed in rare situations, even with collision avoidance constraints due to the physical limitations of the vehicle. It can be observed that the inclusion of the safety cost in any combination increases the safety metrics as observed through the increased distance to pedestrians ($d_{ped}$) and reduced probability of collision ($p_{col}$). Incorporating a safety cost in addition to the collision avoidance constraints helps the vehicle in slowing down earlier, as expressed by the lower average velocity ($a_m$) and higher interaction time ($t_{int}$).

The inclusion of the comfort cost improved the comfort metrics as expected by having a lower absolute jerk ($j_m$) and acceleration values ($a_m$). Similarly, the inclusion of the performance cost reduces the time to destination ($t_{des}$) but increases the chances of collision ($p_{col}$). An interesting observation is that the inclusion of performance cost sometimes results in the vehicle not yielding to the pedestrians expressed by $p_{ny}$. A reason for this could be that the AV velocity is higher in this case than in the other cases, and the pedestrian is close to the AV such that it is safer to accelerate and pass the crosswalk than to decelerate and try to stop.

We explored how the different vehicle objectives affect the vehicle performance during interaction with pedestrians. We found that a safety cost in addition to

Table 6.5: Comparison of average performance metrics for pedestrians for the various weight combinations.

| **Metric** | **S** | **P** | **C** | **PC** | **SC** | **SP** | **SPC** |
|---|---|---|---|---|---|---|---|
| $t_{int}$ $[s]$ | 7.54 | 5.38 | 6.08 | 6.09 | 7.21 | 7.18 | 7.01 |
| $t_{des}$ $[s]$ | 9.75 | 7.46 | 9.47 | 8.73 | 8.78 | 9.08 | 8.91 |
| $d_{ped}$ $[m]$ | 7.34 | 4.33 | 5.20 | 5.85 | 6.38 | 6.48 | 6.21 |
| $v_m$ $[m/s]$ | 12.10 | 13.80 | 13.30 | 13.19 | 12.38 | 12.08 | 12.20 |
| $a_m$ $[m/s^2]$ | -0.15 | -0.20 | -0.06 | -0.19 | -0.14 | -0.11 | -0.12 |
| $j_m$ $[m/s^3]$ | 0.08 | 0.10 | 0.04 | 0.08 | 0.07 | 0.12 | 0.09 |
| $p_{col}$ | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| $p_{ny}$ | 0.00 | 0.05 | 0.02 | 0.03 | 0.00 | 0.00 | 0.00 |

$t_{int}$ - interaction duration, time when AVs first predicted pedestrians to step into the lane to the time when the pedestrians step out of the lane or when AVs crossed the pedestrian, whichever is earlier
$t_{des}$ - time for the AV to reach the destination
$d_{ped}$ - minimum distance to pedestrian when pedestrian is crossing
$v_m$ - average velocity during interaction defined by $t_{int}$
$a_m$ - average acceleration during interaction
$j_m$ - average absolute jerk during interaction
$p_{col}$ - probability of collision given by the ratio of the number of collisions observed to the total number of simulation run for that condition
$p_{ny}$ - probability of not yielding to the pedestrians given by the number of runs the vehicle did not yield to the total number of simulation run for that condition

collision avoidance constraints greatly improves the safety performance metrics. Our findings suggest that AV decision-making can potentially be expressed by adjusting the weights of the different costs. Various weight combinations corresponding to different driving behaviors — cautious, aggressive, nudging, comfortable, etc., — can be determined and adaptively altered based on the dynamic changes in the situation.

## 6.7 Chapter summary

This chapter discusses the development of an AV controller for safe and trustworthy motion planning that incorporates the predictions from the pedestrian behavior models. The controller demonstrates the ability to express AV intent from the optimization weights. The implications of varying AV behavior through the weights include (i) designing AV driving behaviors that can potentially be understood by the pedestrians and (ii) adaptively varying weights based on changes in the context, both of which could facilitate safe and trustworthy navigation.

# CHAPTER VII

# Conclusion

As discussed in Chapter I, public trust in and acceptance of automated vehicles (AVs) are necessary to realize the several benefits of AVs, including improved traffic safety, transportation accessibility, efficiency, and comfort. A major operational challenge for AVs is navigating urban streets filled with pedestrians, the most vulnerable in case of crashes. Demonstrating safe and trustworthy navigation in such complex situations could aid the AVs in gaining public trust and acceptance. Current AV planning methods mostly employ a conservative approach to maximize safety but can restrict traffic flow while eliciting unexpected public reactions ranging from curiosity to vandalism. Understanding and predicting pedestrian behavior can give the AVs sufficient time to plan safe trajectories.

This dissertation characterizes the relationships between pedestrian behavior, AV behavior, and pedestrian trust in the AVs and develops models of pedestrian and AV behaviors for safe and trustworthy navigation. The four primary contributions of this dissertation are stated below.

## 7.1 Contributions

1. *C1: Characterization of how driving behavior and vehicle type affect pedestrian's trust in AVs.*

   The first main contribution of this dissertation is the characterization of the role of driving behavior and vehicle type (HDV or AV) on pedestrian trust and pedestrian behaviors. We conducted two user studies to investigate pedestrian interaction with AVs, which are described in Chapter III and in [33, 34]. The first user study investigated the effects of driving behavior and crosswalk type on pedestrian trust and behavior, and the findings can be summarized as fol-

lows. Similar to current research [18, 19], we found that AV intent can be deduced from their behaviors. Pedestrian trust and behaviors are influenced by AV driving behavior. More trust elicited observable trusting behaviors from the pedestrians, such as reduced gaze at AVs, increased walking speed, and increased waiting time before crossing. However, the influence of AV driving behavior on pedestrian trust and behavior is applicable only at unsignalized crosswalks. The second user study investigated the effects of vehicle type on pedestrian trust, and the findings can be summarized as follows. Pedestrians have different expectations towards AVs than HDVs, resulting in different pedestrian trust for AVs compared to HDVs. Pedestrians' trust in the AVs increases with each subsequent interaction.

The findings from these two user studies were instrumental in developing pedestrian behavior models and AV control algorithms which are the subsequent contributions in this dissertation.

2. *C2: Development of an explainable and computationally efficient long-term model for multimodal pedestrian behavior.*

   The second contribution of this dissertation is the development of an explainable and computationally efficient pedestrian behavior model that is suitable for long-term multimodal predictions. As described in Chapter IV and in [35, 36], a pedestrian behavior model based on hybrid automata theory was developed and validated on two datasets — one collected from the VR study discussed in Section 3.2.3.4 and the other is the inD dataset that captured pedestrian-HDV interactions in the real-world [37]. Compared to baseline methods, the developed model better predicted long-term multimodal pedestrian trajectories.

3. *C3: Extension of pedestrian behavior model incorporating interactions among multiple pedestrians.*

   The third contribution of this dissertation is the development of a method that efficiently incorporates the effects of pedestrian interaction with neighboring pedestrians. As described in Chapter V, the pedestrian modeling framework developed in Chapter IV was extended to incorporate pedestrian interactions with neighboring pedestrians. The extended model was validated on the diverse and large-scale nuScenes dataset containing real-world interactions between pedestrians and vehicles [40]. Further, we identified a limiting interaction region for pedestrians beyond which pedestrian behavior is not significantly affected by

other pedestrians, which could be used to identify relevant neighboring pedestrians by the AVs.

4. *C4: Development of a behavior-aware controller*

   The fourth contribution of this dissertation is the development of an AV controller for safe and trustworthy motion planning that incorporates the predictions from the pedestrian behavior models. The controller demonstrates the ability to design driving behaviors that the pedestrians can potentially understand, facilitating safe and trustworthy navigation. The development of a behavior-aware controller based on MPC framework is discussed in Chapter VI and published in [30, 41].

## 7.2 Limitations

This dissertation has explored modeling pedestrian and control of AV behaviors for safe and trustworthy pedestrian-AV interactions. However, several limitations remain related to all the contributions.

The user studies discussed in Chapter III were conducted in both virtual and real-world environments. A limitation is that the studies explored the interaction between only one pedestrian and one or more approaching vehicles. However, as described in Section 2.4, there are multiple pedestrians and multiple vehicles in the real world. Another limitation is that the studies were conducted either in simulated or controlled real-world environments that do not capture the realistic traffic behavior on urban streets. While several studies suggest similarities in pedestrian behavior in the simulators and the real world [146, 186], reproducing the conditions of real-world driving, particularly the risks involved, is challenging. Thus the conclusions and results presented in this dissertation should be validated on actual AVs operating in real-world traffic conditions.

A major limitation for the pedestrian behavior model is that the model assumes that pedestrians always use the crosswalk to cross. Though this enables the simplification of the problem and is valid in most instances, it does not consider jaywalking pedestrians and may pose serious safety concerns. One way to incorporate jaywalking pedestrians is to consider their behavior as an anomalous behavior for which the AVs can take conservative actions such as stopping. Another limitation is that the pedestrians are assumed to travel at a constant velocity within a particular discrete state. Though constant velocity dynamics are found to be valid in many cases [153], it

could result in producing intersecting trajectory predictions when there are multiple pedestrians in the scene. However, the modeling framework is flexible to accommodate other types of continuous motion models that can avoid collisions, such as social forces models. Integrating these models in the hybrid automata framework to produce collision-free trajectories can be a valuable direction for future work. Another limitation is that the models do not consider contextual elements such as traffic signs, signals, etc., or individual attributes such as age, gender, etc. While these factors can improve the behavior predictions, they also increase model complexity as obtaining and processing such data can be computationally expensive.

The behavior-aware model predictive controller (B-MPC) was validated on a simple interaction scenario where the vehicle calculated only its longitudinal dynamics. More complex scenarios involving intersections, curvy roads, etc., are more representative of real-world conditions. Another limitation is that the controller used the most probable predictions from the pedestrian behavior model. This could be extended to include probabilistic predictions using techniques such as chance-constrained optimization.

The B-MPC was safer than the baseline controller as it had higher gap times and distance to pedestrians than the baseline controller. While this can improve a pedestrian's perception of AV safety, the AVs can still collide with the pedestrians due to unexpected pedestrian behaviors (dynamically changing environment). A recent approach is to identify trajectories that prove that the AVs did not take actions that could have potentially resulted in a collision, i.e., they are not at fault even in the case of a collision [10, 187]. While perceived safety primarily affects pedestrian behaviors and response to AVs, the AVs should demonstrate provably safe not-at-fault behaviors to minimize collisions and further improve safety. This, however, is currently not addressed in this research and is a limitation.

## 7.3 Future work

While several limitations and research questions can be identified for the main contributions of this dissertation, three potential research directions that can leverage the work in this dissertation are described in this section.

### 7.3.1 Anomalous pedestrian behavior detection

In this dissertation, common and rational pedestrian behavior was modeled. However, in the real world, not all pedestrians engage in rational behaviors. Some pedes-

trians may engage in anomalous and potentially dangerous behaviors such as jaywalking, thereby reducing traffic safety. Early identification of anomalous behaviors thus allows the vehicles to take corresponding actions that could significantly improve traffic safety.

Using the prediction models we developed in this project, we could develop methods to identify anomalous pedestrian behaviors. Anomalies have been identified in pedestrian motion by observing deviations from the continuous predicted trajectories [74, 188]. We could explore anomalies in discrete, continuous, and a combination of discrete and continuous states. For example, a discrete state transition directly from *Approach* to *Cross* could indicate a jaywalking pedestrian, and a transition from *Cross* to *Wait* or *Approach* could indicate pedestrians who are in and out of the road confusing the oncoming vehicles. More commonly, a comparison of the predicted and actual trajectories would give a measure of deviation. This deviation could occur due to two reasons — the model uncertainty and anomalous pedestrian behavior. The model uncertainty should result in a lower deviation than an anomalous pedestrian behavior, and thus the deviation can be used to identify anomalous behaviors.

### 7.3.2 Formalizing human-robot interaction

Although the models developed in this dissertation have improved performance, they do not come with formal guarantees. People are more likely to trust and deploy AV controllers that are formally verified and provide some safety guarantees. Recently, researchers are focusing on formalizing human-robot interaction (HRI) [189, 190]. Formalizing AV interaction with pedestrians and other road agents can enable the creation of trustworthy AVs and support explicit reasoning about their operation domain, and the context of guarantees [190]. The hybrid automaton structure of the pedestrian behavior models enables formalizing pedestrian behaviors through temporal logic specifications.

A similar hybrid automaton framework can be developed for the AVs and other vehicles and cyclists and explore the interactions between multiple AVs and multiple kinds of human road agents. The framework enables representing AV intent as a sequence of discrete states or linear temporal logic specifications. Further, AV safety verification can be done using model checking, reachability analysis tools, etc. A major implication of formalizing pedestrian-AV interactions is the ability to provably verify safe AV trajectories enabling not-at-fault driving by the AVs.

### 7.3.3 Trust-based decision making

The first contribution of this dissertation characterized the relation between AV driving behavior, crosswalk type, and pedestrians' trust and trusting behaviors. These relationships could be further explored under different situations varying in the amount of risks similar to real-world situations. Trust between agents is required for smooth coordination. For example, when multiple safe trajectories are possible for the AV, choosing the trajectory that might increase trust can improve coordination between the road agents. Pedestrian trusting behaviors, such as their gaze at vehicles, can be used to estimate the trust of the pedestrians in real time. The estimated trust could be used in an optimization-based decision-making framework to choose AV behaviors that are both safe and trustworthy.

## 7.4 Outlook and impact

This dissertation developed computationally efficient pedestrian behavior models and an AV planning method suitable for real-time planning. These models and methods extend the knowledge on safe and trustworthy interaction between humans and robots, particularly applicable to pedestrians and AVs. The relationships characterized in Chapter III between pedestrian trust in AVs, pedestrian trusting behaviors, and several behavioral and environmental factors will aid in choosing AV behaviors that could improve trust.

The multimodal predictions obtained from the pedestrian model (discussed in Chapters IV and V) are driven by the pedestrian's intents and decisions. This will enable the AVs to calculate trajectories that avoid potential collisions and reason about the intents of the pedestrians while taking into account the AV's intent. The predictions can also serve as a means to identify anomalous pedestrian behavior, thereby avoiding any potentially harmful interactions. Though the models require map information, the information is encoded to easily be transferable to new environments with similar map elements (crosswalks, sidewalks, and lanes), making the models generalizable. Because of this, the models would require minimum re-training when deploying to new environments.

The AV planning method will enable AVs to plan for long durations preemptively. Thereby, the resulting AV behavior will not be overly conservative. Further, the AV will have sufficient time to choose paths that are not only safe but also potentially trustworthy. In addition to improving safety, the proposed approach would improve traffic flow as highly conservative driving behavior is unpredictable and disrupts traf-

fic [191]. In addition to reducing fatality, the proposed research through accurate road agent behavior predictions and efficient interactive AV planning could directly improve the overall safety of transportation by reducing the number of non-fatal crashes. This makes commuting more productive and comfortable and indirectly improves the accessibility of transportation to everyone, promoting the acceptance and development of automated vehicles.

# APPENDIX

# APPENDIX A

# User Study 1

## A.1 Latin Square Design

The standard Latin square design we employed in the study is given below, with the six treatment conditions Defensive unsignalized (A), Normal unsignalized (B), Aggressive unsignalized (C), Defensive signalized (D), Normal signalized (E), and Aggressive signalized (F). Each condition appears exactly once in each row and once in each column, which resulted in a set of six condition sequences. The set was designed such that every treatment condition appears exactly once before and once after every other condition. The set was repeated five times to get thirty condition sequences for the thirty participants in the study.

| | **Order of treatment conditions** | | | | | |
|---|---|---|---|---|---|---|
| **Subject Number** | **1st** | **2nd** | **3rd** | **4th** | **5th** | **6th** |
| **1** | A | F | B | E | C | D |
| **2** | B | A | C | F | D | E |
| **3** | C | B | D | A | E | F |
| **4** | D | C | E | B | F | A |
| **5** | E | D | F | C | A | B |
| **6** | F | E | A | D | B | C |

## A.2 Post-Treatment Trust Questionnaire

The below questionnaire has been adapted from [133], which examines trust in automation.

Please indicate the extent to which you believe the autonomy has each of the following traits (from 1 representing "none at all" to 7 representing "extremely high").

1. Competence: To what extent did the autonomous cars perform their function properly i.e. recognizing you and reacting for you?
2. Predictability: To what extent were you able to predict the behavior of the autonomous cars from moment to moment?
3. Dependability: To what extent can you count on the autonomous cars to do its job?
4. Responsibility: To what extent the autonomous cars seemed to be wary of their surroundings?
5. Reliability over time: To what extent do you think the autonomous car's actions were consistent throughout the interaction?
6. Faith: What degree of faith do you have that the autonomous cars will be able to cope with all uncertainties in the future?

## A.3 Simulator Sickness Questionnaire (SSQ)

Please indicate how much each symptom below is affecting you right now (survey from [134]).

| **Sensation** | **0** | **1** | **2** | **3** |
|---|---|---|---|---|
| General Discomfort | None | Slight | Moderate | Severe |
| Fatigue | None | Slight | Moderate | Severe |
| Headache | None | Slight | Moderate | Severe |
| Eye Strain | None | Slight | Moderate | Severe |
| Difficulty focusing | None | Slight | Moderate | Severe |
| Increased salivation | None | Slight | Moderate | Severe |
| Sweating | None | Slight | Moderate | Severe |
| Nausea | None | Slight | Moderate | Severe |
| Difficulty concentrating | None | Slight | Moderate | Severe |
| Fullness of head | None | Slight | Moderate | Severe |
| Blurred vision | None | Slight | Moderate | Severe |
| Dizzy (eyes open) | None | Slight | Moderate | Severe |
| Dizzy (eyes closed) | None | Slight | Moderate | Severe |
| Vertigo | None | Slight | Moderate | Severe |
| Stomach awareness | None | Slight | Moderate | Severe |
| Burping | None | Slight | Moderate | Severe |

# BIBLIOGRAPHY


[1] "Power companies build for your new electric living." In: *The Victoria Advocate* (Mar. 1956).

[2] Dean A Pomerleau. *Alvinn: An autonomous land vehicle in a neural network*. Tech. rep. Carnegie Mellon University, 1989.

[3] Todd Litman. *Autonomous Vehicle Implementation Predictions*. Victoria, Canada: Victoria Transport Policy Institute, 2019.

[4] *NHTSA: Automated Vehicles for Safety*. `https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety`. (Accessed on 05/02/2021).

[5] Daniel J Fagnant and Kara Kockelman. "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations". In: *Transportation Research Part A: Policy and Practice* 77 (2015), pp. 167–181.

[6] Peng Liu, Run Yang, and Zhigang Xu. "Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions". In: *Risk Analysis* 39.2 (2019), pp. 326–341.

[7] Peng Jing, Gang Xu, Yuexia Chen, Yuji Shi, and Fengping Zhan. "The determinants behind the acceptance of autonomous vehicles: a systematic review". In: *Sustainability* 12.5 (2020), p. 1719.

[8] Neal E. Boudette. *Tesla's Autopilot Technology Faces Fresh Scrutiny*. Mar. 2021. URL: `https://www.nytimes.com/2021/03/23/business/teslas-autopilot-safety-investigations.html`.

[9] *Uber's self-driving operator charged over fatal crash*. Sept. 2020. URL: `https://www.bbc.com/news/technology-54175359`.

[10] Andrea Bajcsy, Somil Bansal, Eli Bronstein, Varun Tolani, and Claire J Tomlin. "An efficient reachability-based framework for provably safe autonomous navigation in unknown environments". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 1758–1765.

[11] Shreyas Kousik, Bohao Zhang, Pengcheng Zhao, and Ram Vasudevan. "Safe, Optimal, Real-time Trajectory Planning with a Parallel Constrained Bernstein Algorithm". In: *IEEE Transactions on Robotics* (2020), pp. 1–6.

[12] Miltos Kyriakidis, Riender Happee, and Joost CF de Winter. "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents". In: *Transportation research part F: traffic psychology and behaviour* 32 (2015), pp. 127–140.

[13] SAE-International. "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles". In: (2020).

[14] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. "Human motion trajectory prediction: A survey". In: *The International Journal of Robotics Research* 39.8 (2020), pp. 895–935.

[15] Dylan Moore, Rebecca Currano, Michael Shanks, and David Sirkin. "Defense against the dark cars: Design principles for griefing of autonomous vehicles". In: *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 2020, pp. 201–209.

[16] Amir Rasouli and John K Tsotsos. "Autonomous vehicles that interact with pedestrians: A survey of theory and practice". In: *IEEE Transactions on Intelligent Transportation Systems* 21 (3 2019), pp. 900–918.

[17] Athanasios Theofilatos, Apostolos Ziakopoulos, Oscar Oviedo-Trespalacios, and Andrew Timmis. "To cross or not to cross? Review and meta-analysis of pedestrian gap acceptance decisions at midblock street crossings". In: *Journal of Transport & Health* 22 (2021), p. 101108.

[18] Claudia Ackermann, Matthias Beggiato, Luka-Franziska Bluhm, and Josef Krems. "Vehicle movement and its potential as implicit communication signal for pedestrians and automated vehicles". In: *Proceedings of the 6th Humanist Conference*. HUMANIST Publications The Hague. 2018, p. 7.

[19] Tanja Fuest, Lars Michalowski, Luca Träris, Hanna Bellem, and Klaus Bengler. "Using the driving behavior of an automated vehicle to communicate intentions-a wizard of oz study". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2018, pp. 3596–3601.

[20] Richard M Sorrentino, John G Holmes, Steven E Hanna, and Ann Sharp. "Uncertainty orientation and trust in close relationships: individual differences in cognitive styles." In: *Journal of Personality and Social Psychology* 68.2 (1995), p. 314.

[21] Julian F.P. Kooij, Fabian Flohr, Ewoud A.I. Pool, and Dariu M. Gavrila. "Context-Based Path Prediction for Targets with Switching Dynamics". In: *International Journal of Computer Vision* 127.3 (2019), pp. 239–262. ISSN: 15731405. DOI: 10.1007/s11263-018-1104-4.

[22] Raul Quintero Minguez, Ignacio Parra Alonso, David Fernandez-Llorca, and Miguel Angel Sotelo. "Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition". In: *IEEE Transactions on Intelligent Transportation Systems* 20.5 (2018), pp. 1803–1814. DOI: 10.1109/TITS.2018.2836305.

[23] Christoph G. Keller and Dariu M. Gavrila. "Will the pedestrian cross? A study on pedestrian path prediction". In: *IEEE Transactions on Intelligent Transportation Systems* 15.2 (2014), pp. 494–506. DOI: 10.1109/TITS.2013.2280766.

[24] Stuart Reid. *DfT Shared Space Project Stage 1: Appraisal of Shared Space*. Tech. rep. 2009.

[25] Khaled Saleh, Mohammed Hossny, and Saeid Nahavandi. "Towards trusted autonomous vehicles from vulnerable road users perspective". In: *2017 11th Annual IEEE International Systems Conference, SysCon*. 2017, pp. 1–7. ISBN: 9781509046225. DOI: 10.1109/SYSCON.2017.7934782.

[26] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezatofighi, and Silvio Savarese. "Social-BiGAT: Multimodal Trajectory Forecasting using Bicycle-GAN and Graph Attention Networks". In: *Advances in Neural Information Processing Systems*. 2019, pp. 137–146.

[27] Charlie Tang and Russ R Salakhutdinov. "Multiple futures prediction". In: *Advances in Neural Information Processing Systems*. 2019, pp. 15424–15434.

[28] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data". In: *arXiv preprint arXiv:2001.03093* (2020).

[29] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. "Planning and decision-making for autonomous vehicles". In: *Annual Review of Control, Robotics, and Autonomous Systems* 1 (2018), pp. 187–210.

[30] S. K. Jayaraman, L. P. Robert, X. J. Yang, A. K. Pradhan, and D. M. Tilbury. "Efficient Behavior-aware Control of Automated Vehicles at Crosswalks using Minimal Information Pedestrian Prediction Model". In: *2020 American Control Conference (ACC)*. 2020, pp. 4362–4368.

[31] Przemyslaw A Lasota, Terrence Fong, and Julie A Shah. "A survey of methods for safe human-robot interaction". In: *Foundations and Trends in Robotics* 5.4 (2014), pp. 261–349.

[32] Douglas Shinkle. "Pedestrian Crossing: 50 State Summary". In: *Proc. Nat. Conf. State Legislatures* (2016). URL: http://www.%20ncsl.org/Res./Transp./%20pedestrian-%20crossing-50-%20state-summary.aspx.

[33] Suresh Kumaar Jayaraman, Chandler Creech, Lionel P. Robert Jr., Dawn M. Tilbury, X. Jessie Yang, Anuj K. Pradhan, and Katherine M. Tsui. "Trust in AV: An Uncertainty Reduction Model of AV-Pedestrian Interactions". In: *Companion 2018 ACM/IEEE Int. Conf. Human-Robot Interaction*. 2018, pp. 133–134. ISBN: 9781450356152. DOI: 10.1145/3173386.3177073.

[34] Suresh Jayaraman, Chandler Creech, Tilbury Dawn, X Jessie Yang, Anuj Pradhan, Katherine Tsui, and Lionel Robert. "Pedestrian Trust in Automated Vehicles: Role of Traffic Signal and AV Driving Behavior". In: *Frontiers in Robotics and AI* 6 (1 2019).

[35] Suresh Kumaar Jayaraman, Dawn Tilbury, Jessie Yang, Anuj Pradhan, and Lionel Robert. "Analysis and Prediction of Pedestrian Crosswalk Behavior during Automated Vehicle Interactions". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 2020.

[36] Suresh Kumaar Jayaraman, Lionel P Robert, X Jessie Yang, and Dawn M Tilbury. "Multimodal Hybrid Pedestrian: A Hybrid Automaton Model of Urban Pedestrian Behavior for Automated Driving Applications". In: *IEEE Access* 9 (2021), pp. 27708–27722.

[37] Julian Bock, Robert Krajewski, Tobias Moers, Steffen Runde, Lennart Vater, and Lutz Eckstein. "The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2020, pp. 1929–1934.

[38] Cyrus Anderson, Ram Vasudevan, and Matthew Johnson-Roberson. "Off the beaten sidewalk: Pedestrian prediction in shared spaces for autonomous vehicles". In: *IEEE Robotics and Automation Letters* 5.4 (2020), pp. 6892–6899.

[39] Jiachen Li, Fan Yang, Masayoshi Tomizuka, and Chiho Choi. "EvolveGraph: Heterogeneous Multi-Agent Multi-Modal Trajectory Prediction with Evolving Interaction Graphs". In: *arXiv preprint arXiv:2003.13924* (2020).

[40] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. "nuscenes: A multimodal dataset for autonomous driving". In: *arXiv preprint arXiv:1903.11027* (2019).

[41] Suresh Kumaar Jayaraman, Lionel P Robert, X Jessie Yang, and Dawn M Tilbury. "Automated Vehicle Behavior Design for Pedestrian Interactions at Unsignalized Crosswalks". In: *International Symposium on Transportation Data and Modelling* (2021).

[42] Charles M DiPietro and L Ellis King. "Pedestrian gap-acceptance". In: *Highway Research Record* 308 (1970).

[43] G. Yannis, E. Papadimitriou, and A. Theofilatos. "Pedestrian gap acceptance for mid-block street crossing". In: *Transportation Planning and Technology* 36.5 (2013), pp. 450–462. ISSN: 03081060. DOI: 10.1080/03081060.2013.818274.

[44] B. Raghuram Kadali, P. Vedagiri, and Nivedan Rathi. "Models for pedestrian gap acceptance behaviour Anal. at unprotected mid-block crosswalks under mixed traffic conditions". In: *Transportation Research Part F: Traffic Psychology Behaviour* 32 (2015), pp. 114–126. DOI: 10.1016/j.trf.2015.05.006.

[45] Marcus A Brewer, Kay Fitzpatrick, Jeffrey A Whitacre, and Dominique Lord. "Exploration of pedestrian gap-acceptance behavior at selected locations". In: *Transportation Research Record* 1982.1 (2006), pp. 132–140.

[46] Zhijie Fang and Antonio M. López. "Is the Pedestrian going to Cross? Answering by 2D Pose Estimation". In: *2018 IEEE Intelligent Vehicles Symposium*. 2018, pp. 1271–1276. arXiv: 1807.10580.

[47] Benjamin Volz, Karsten Behrendt, Holger Mielenz, Igor Gilitschenski, Roland Siegwart, and Juan Nieto. "A data-driven approach for pedestrian intention estimation". In: *IEEE International Conference on Intelligent Transportation Systems*. 2016, pp. 2607–2612.

[48] Matus Sucha, Daniel Dostal, and Ralf Risser. "Pedestrian-driver communication and decision strategies at marked crossings". In: *Accident Anal. Prevention* (2017). DOI: 10.1016/j.aap.2017.02.018.

[49] Nicolas Guéguen, Sébastien Meineri, and Chloé Eyssartier. "A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing". In: *Safety science* 75 (2015), pp. 87–89.

[50] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. "Understanding pedestrian behavior in complex traffic scenes". In: *IEEE Transactions on Intelligent Vehicles* 3.1 (2017), pp. 61–70.

[51] Natasha Merat, Tyron Louw, Ruth Madigan, Marc Wilbrink, and Anna Schieben. "What externally presented information do VRUs require when interacting with fully Automated Road Transport Systems in shared space?" In: *Accident Analysis & Prevention* 118 (2018), pp. 244–252.

[52] Samantha Reig, Selena Norman, Cecilia G Morales, Samadrita Das, Aaron Steinfeld, and Jodi Forlizzi. "A field study of pedestrians and autonomous vehicles". In: *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 2018, pp. 198–209.

[53] Chia-Ming Chang, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi. "Eyes on a Car: an Interface Design for Communication between an Autonomous Car and a Pedestrian". In: *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. 2017, pp. 65–73.

[54] Chia-Ming Chang, Koki Toda, Takeo Igarashi, Masahiro Miyata, and Yasuhiro Kobayashi. "A video-based study comparing communication modalities between an autonomous car and a pedestrian". In: *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 2018, pp. 104–109.

[55] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, and Stas Krupenia. "Communicating intent of automated vehicles to pedestrians". In: *Frontiers in psychology* 9 (2018), pp. 1336–1352.

[56] Karthik Mahadevan. "Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction". In: *Proc. 2018 CHI Conf. Human Factors Computing Syst.* 2018, p. 429.

[57] Debargha Dey and Jacques Terken. "Pedestrian interaction with vehicles: roles of explicit and implicit communication". In: *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. 2017, pp. 109–113.

[58] Anantha Pillai. "Virtual reality based study to analyse pedestrian attitude towards autonomous vehicles". MA thesis. KTH Royal Institute of Technology, 2017, p. 14.

[59] Henri Schmidt, Jack Terwilliger, Dina AlAdawy, and Lex Fridman. "Hacking nonverbal communication between pedestrians and vehicles in virtual reality". In: *arXiv preprint arXiv:1904.01931* (2019).

[60] Mark Meeder, Ernst Bosina, and Ulrich Weidmann. "Autonomous vehicles: Pedestrian heaven or pedestrian hell". In: *17th Swiss Transport Research Conference*. 2017, pp. 17–19.

[61] Adam Millard-Ball. "Pedestrians, autonomous vehicles, and cities". In: *Journal of planning education and research* 38.1 (2018), pp. 6–12.

[62] Victor Malmsten Lundgren, Azra Habibovic, Jonas Andersson, Tobias Lagström, Maria Nilsson, Anna Sirkka, Johan Fagerlönn, Rikard Fredriksson, Claes Edgren, and Stas Krupenia. "Will there be new communication needs when introducing automated vehicles to the urban context?" In: *Advances in human aspects of transportation*. Springer, 2017, pp. 485–497.

[63] Dirk Rothenbucher, Jamy Li, David Sirkin, Brian Mok, and Wendy Ju. "Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles". In: *25th IEEE Int. Symp. Robot Human Interactive Commun., RO-MAN 2016*. 2016, pp. 795–802. ISBN: 9781509039296. DOI: 10.1109/ROMAN.2016.7745210.

[64] Raphael Zimmermann and Reto Wettach. "First step into visceral interaction with autonomous vehicles". In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 2017, pp. 58–64.

[65] John D Lee and Neville Moray. "Trust, self-confidence, and operators' adaptation to automation". In: *International journal of human-computer studies* 40.1 (1994), pp. 153–184.

[66] John D Lee and Katrina A See. "Trust in automation: Designing for appropriate reliance". In: *Human factors* 46.1 (2004), pp. 50–80.

[67] Victor Riley. "Operator reliance on automation: Theory and data". In: *Automation and human performance: Theory and applications* (1996), pp. 19–35.

[68] Ana Rodríguez Palmeiro, Sander van der Kint, Luuk Vissers, Haneen Farah, Joost C.F. de Winter, and Marjan Hagenzieker. "Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment". In: *Transp. Res. Part F: Traffic Psychology Behaviour* 58 (2018), pp. 1005–1020. ISSN: 13698478. DOI: 10.1016/j.trf.2018.07.020.

[69] Andreas Møgelmose, Mohan M Trivedi, and Thomas B Moeslund. "Trajectory Analysis and Prediction for Improved Pedestrian Safety: Integrated Framework and Evaluations". In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2015, pp. 330–335.

[70] Nicolas Schneider and Dariu M Gavrila. "Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study". In: *German Conference on Pattern Recognition*. Springer. 2013, pp. 174–183.

[71] Rohan Chandra, Uttaran Bhattacharya, Aniket Bera, and Dinesh Manocha. "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2019, pp. 8483–8492.

[72] Nachiket Deo and Mohan M Trivedi. "Multi-modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs". In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2018, pp. 1179–1184.

[73] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 2255–2264.

[74] Yutao Han, Rina Tse, and Mark Campbell. "Pedestrian Motion Model using Non-parametric Trajectory Clustering and Discrete Transition Points". In: *IEEE Robotics and Automation Letters* 4.3 (2019), pp. 2614–2621.

[75] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 1349–1358.

[76] Anirudh Vemula, Katharina Muelling, and Jean Oh. "Modeling cooperative navigation in dense human crowds". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2017, pp. 1685–1692.

[77] Vasiliy Karasev and Stefano Soatto. "Intent-Aware Long-Term Prediction of Pedestrian Motion". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 2016. ISBN: 9781467380263.

[78] Kris M Kitani, Brian D Ziebart, J Andrew Bagnell, and Martial Hebert. "Activity Forecasting". In: *European Conference on Computer Vision*. Springer, 2012, pp. 201–214.

[79] Andrey Rudenko, Luigi Palmieri, and Kai O Arras. "Joint long-term prediction of human motion using a planning-based social force approach". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2018, pp. 1–7.

[80] Brian D Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J Andrew Bagnell, Martial Hebert, Anind K Dey, and Siddhartha Srinivasa. "Planning-based prediction for pedestrians". In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2009, pp. 3931–3936.

[81] Nachiket Deo and Mohan M Trivedi. "Learning and Predicting on-road Pedestrian Behavior Around Vehicles". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 1–6.

[82] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio 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.

[83] Yuexin Ma, Xinge Zhu, Sibo Zhang, Ruigang Yang, Wenping Wang, and Dinesh Manocha. "TrafficPredict: Trajectory Prediction for Heterogeneous Traffic-Agents". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 6120–6127.

[84] Nicholas Rhinehart, Rowan McAllister, Kris Kitani, and Sergey Levine. "PRECOG: Prediction Conditioned on Goals in Visual Multi-Agent Settings". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 2821–2830.

[85] Federico Bartoli, Giuseppe Lisanti, Lamberto Ballan, and Alberto Del Bimbo. "Context-aware trajectory prediction". In: *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE. 2018, pp. 1941–1946.

[86] Boris Ivanovic and Marco Pavone. "The Trajectron: Probabilistic Multi-Agent Trajectory Modeling with Dynamic Spatiotemporal Graphs". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 2375–2384.

[87] Kevin Patrick Murphy. "Dynamic bayesian networks: representation, inference and learning". PhD thesis. University of California, Berkeley Berkeley, CA, 2002.

[88] Petter Nilsson, Omar Hussien, Ayca Balkan, Yuxiao Chen, Aaron D Ames, Jessy W Grizzle, Necmiye Ozay, Huei Peng, and Paulo Tabuada. "Correct-by-construction adaptive cruise control: Two approaches". In: *IEEE Transactions on Control Systems Technology* 24.4 (2015), pp. 1294–1307.

[89] Niccoló Bisagno, Bo Zhang, and Nicola Conci. "Group LSTM: Group trajectory prediction in crowded scenarios". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 213–225.

[90] Andreas T Schulz and Rainer Stiefelhagen. "A Controlled Interactive Multiple Model Filter for Combined Pedestrian Intention Recognition and Path Prediction". In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE. 2015, pp. 173–178.

[91] Florian Kuhnt, Jens Schulz, Thomas Schamm, and J Marius Zöllner. "Understanding interactions between traffic participants based on learned behaviors". In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2016, pp. 1271–1278.

[92] Gabriel Agamennoni, Juan I Nieto, and Eduardo M Nebot. "Estimation of Multivehicle Dynamics by Considering Contextual Information". In: *IEEE Transactions on Robotics* 28.4 (2012), pp. 855–870.

[93] Tobias Gindele, Sebastian Brechtel, and Rüdiger Dillmann. "A Probabilistic Model for Estimating Driver Behaviors and Vehicle Trajectories in Traffic Environments". In: *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE. 2010, pp. 1625–1631.

[94] Claudia Blaiotta. "Learning generative socially aware models of pedestrian motion". In: *IEEE Robotics and Automation Letters* 4.4 (2019), pp. 3433–3440.

[95] Chiho Choi, Abhishek Patil, and Srikanth Malla. "DROGON: A Causal Reasoning Framework for Future Trajectory Forecast". In: *arXiv preprint arXiv:1908.00024* (2019).

[96] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher B Choy, Philip H S Torr, and Manmohan Chandraker. "DESIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 336–345.

[97] Junwei Liang, Lu Jiang, Kevin Murphy, Ting Yu, and Alexander Hauptmann. "The garden of forking paths: Towards multi-future trajectory prediction". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 10508–10518.

[98] Nachiket Deo and Mohan M Trivedi. "Convolutional Social Pooling for Vehicle Trajectory Prediction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 1468–1476.

[99] Sergey Prokudin, Peter Gehler, and Sebastian Nowozin. "Deep directional statistics: Pose estimation with uncertainty quantification". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 534–551.

[100] Osama Makansi, Eddy Ilg, Ozgun Cicek, and Thomas Brox. "Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7144–7153.

[101] Matthias Althoff, Daniel Heß, and Florian Gambert. "Road Occupancy Prediction of Traffic Participants". In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE. 2013, pp. 99–105.

[102] Yanlei Gu, Yoriyoshi Hashimoto, Li-Ta Hsu, and Shunsuke Kamijo. "Motion planning based on learning models of pedestrian and driver behaviors". In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2016, pp. 808–813.

[103] Aniket Bera, Sujeong Kim, Tanmay Randhavane, Srihari Pratapa, and Dinesh Manocha. "GLMP-realtime Pedestrian Path Prediction using Global and Local Movement Patterns". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2016, pp. 5528–5535.

[104] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. "Vision meets Robotics: The KITTI Dataset". In: *International Journal of Robotics Research (IJRR)* 32 (11 2013), pp. 1231–1237.

[105] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. "Learning social etiquette: Human trajectory understanding in crowded scenes". In: *Proceedings of the European conference on computer vision (ECCV)*. Springer. 2016, pp. 549–565.

[106] Ivo Batkovic, Mario Zanon, Nils Lubbe, and Paolo Falcone. "A computationally efficient model for pedestrian motion prediction". In: *2018 European Control Conference (ECC)*. IEEE. 2018, pp. 374–379.

[107] Sean Vaskov, Utkarsh Sharma, Shreyas Kousik, Matthew Johnson-Roberson, and Ramanarayan Vasudevan. "Guaranteed safe reachability-based trajectory design for a high-fidelity model of an autonomous passenger vehicle". In: *2019 American Control Conference (ACC)*. IEEE. 2019, pp. 705–710.

[108] Jaime F Fisac, Eli Bronstein, Elis Stefansson, Dorsa Sadigh, S Shankar Sastry, and Anca D Dragan. "Hierarchical game-theoretic planning for autonomous vehicles". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 9590–9596.

[109] Nan Li, Dave W Oyler, Mengxuan Zhang, Yildiray Yildiz, Ilya Kolmanovsky, and Anouck R Girard. "Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems". In: *IEEE Transactions on control systems technology* 26.5 (2017), pp. 1782–1797.

[110] Dorsa Sadigh, Nick Landolfi, Shankar S Sastry, Sanjit A Seshia, and Anca D Dragan. "Planning for cars that coordinate with people: leveraging effects on human actions for planning and active information gathering over human internal state". In: *Autonomous Robots* 42.7 (2018), pp. 1405–1426.

[111] Sisi Li, Nan Li, Anouck Girard, and Ilya 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.

[112] Matthias Althoff and John M Dolan. "Online verification of automated road vehicles using reachability analysis". In: *IEEE Transactions on Robotics* 30.4 (2014), pp. 903–918.

[113] David Fridovich-Keil, Sylvia L Herbert, Jaime F Fisac, Sampada Deglurkar, and Claire J Tomlin. "Planning, fast and slow: A framework for adaptive real-time safe trajectory planning". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2018, pp. 387–394.

[114] Anirudha Majumdar and Russ Tedrake. "Funnel libraries for real-time robust feedback motion planning". In: *The International Journal of Robotics Research* 36.8 (2017), pp. 947–982.

[115] Leslie A Baxter and Barbara M Montgomery. *Relating: Dialogues and Dialectics*. New York, NY: Guilford Press, 1996.

[116] Michael W Kramer. "Motivation to reduce uncertainty: A reconceptualization of uncertainty reduction theory". In: *Management Communication Quarterly* 13.2 (1999), pp. 305–316.

[117] Michael Sunnafrank. "Predicted Outcome Value during Initial Interactions: A Reformulation of Uncertainty Reduction Theory". In: *Human Communication Research* 13.1 (1986), pp. 3–33.

[118] Jason A Colquitt, Jeffery A LePine, Ronald F Piccolo, Cindy P Zapata, and Bruce L Rich. "Explaining the Justice–Performance Relationship: Trust as Exchange Deepener or Trust as Uncertainty Reducer?" In: *Journal of Applied Psychology* 97.1 (2012), p. 1.

[119] J David Lewis and Andrew Weigert. "Trust as a Social Reality". In: *Social Forces* 63.4 (1985), pp. 967–985.

[120] Lionel P Robert, Alan R Denis, and Yu-Ting Caisy Hung. "Individual swift trust and knowledge-based trust in face-to-face and virtual team members". In: *Journal of Management Information Systems* 26.2 (2009), pp. 241–279.

[121] Tove Helldin, Göran Falkman, Maria Riveiro, and Staffan Davidsson. "Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving". In: *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2013), pp. 210–217.

[122] Louis Mizell, Matthew Joint, and Dominic Connell. "Aggressive driving: Three studies". In: *AAA Foundation for Traffic Safety* (1997), pp. 1–13.

[123] Friederike Schneemann and Irene Gohl. "Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments". In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2016, pp. 38–43.

[124] Seiji SC Steimetz. "Defensive Driving and the External Costs of Accidents and Travel Delays". In: *Transportation Research Part B: Methodological* 42.9 (2008), pp. 703–724.

[125] Ilja T Feldstein, Christian Lehsing, André Dietrich, and Klaus Bengler. "Pedestrian simulators for traffic research: State of the art and future of a motion lab". In: *International Journal of Human Factors Modelling and Simulation* 6.4 (2018), pp. 250–265.

[126] Viola Cavallo, Aurélie Dommes, Nguyen-Thong Dang, and Fabrice Vienne. "A street-crossing simulator for studying and training pedestrians". In: *Transportation research part F: traffic psychology and behaviour* 61 (2019), pp. 217–228.

[127] Shuchisnigdha Deb, Daniel W Carruth, Richard Sween, Lesley Strawderman, and Teena M Garrison. "Efficacy of virtual reality in pedestrian safety research". In: *Applied Ergonomics* 65 (2017), pp. 449–460.

[128] Karthik Mahadevan, Elaheh Sanoubari, Sowmya Somanath, James E Young, and Ehud Sharlin. "AV-Pedestrian interaction design using a pedestrian mixed traffic simulator". In: *Proceedings of the 2019 on designing interactive systems conference*. 2019, pp. 475–486.

[129] Cynthia G Baum, Rex Forehand, and Leslie E Zegiob. "A Review of Observer Reactivity in Adult-Child Interactions". In: *Journal of Behavioral Assessment* 1.2 (1979), pp. 167–178.

[130] Tom Urbanik, Alison Tanaka, Bailey Lozner, Eric Lindstrom, Kevin Lee, Shaun Quayle, Scott Beaird, Shing Tsoi, Paul Ryus, and Doug Gettman. *Signal Timing Manual*. 2nd ed. Vol. 1. Washington, D.C: Transportation Research Board, 2015.

[131] James R Lewis. "Pairs of latin squares to counterbalance sequential effects and pairing of conditions and stimuli". In: *Proceedings of the Human Factors Society Annual Meeting*. Vol. 33. 18. SAGE Publications Sage CA: Los Angeles, CA. 1989, pp. 1223–1227.

[132] Kimberly C Preusse and Wendy A Rogers. "Error Interpretation during Everyday Automation Use". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60.1 (2016), pp. 805–809.

[133] B.M. Muir. "Trust between humans and machines, and the design of decision aids". In: *International Journal of Man-Machine Studies* 27(5-6) (1987), pp. 527–539.

[134] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness". In: *The International Journal of Aviation Psychology* 3.3 (1993), pp. 203–220.

[135] Sebastian Hergeth, Lutz Lorenz, Roman Vilimek, and Josef F Krems. "Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving". In: *Human Factors* 58.3 (2016), pp. 509–519.

[136] Walter W Stroup. *Generalized linear mixed models: modern concepts, methods and applications*. CRC press, 2012.

[137] Claes Fornell and David F Larcker. "Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics". In: *Journal of Marketing Research* 18.3 (1981), pp. 382–388.

[138] Herbert Glejser. "A New Test for Heteroskedasticity". In: *Journal of the American Statistical Association* 64.325 (1969), pp. 316–323.

[139] M Stone. "Comments on model selection criteria of Akaike and Schwarz". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1979), pp. 276–278.

[140] Lesa Hoffman and Michael J Rovine. "Multilevel models for the experimental psychologist: Foundations and illustrative examples". In: *Behavior Research Methods* 39.1 (2007), pp. 101–117.

[141] Paul D Bliese. "Within-group Agreement, Non-independence, and Reliability: Implications for Data Aggregation and Analysis". In: *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions.* Ed. by Katherine J. Klein and Steve W. J. Kozlowski. San Francisco, CA: Jossey-Bass, 2000, pp. 349–381.

[142] Gowri Asaithambi, Manu O Kuttan, and Sarath Chandra. "Pedestrian road crossing behavior under mixed traffic conditions: A comparative study of an intersection before and after implementing control measures". In: *Transportation in developing economies* 2.2 (2016), p. 14.

[143] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. "Agreeing to cross: How drivers and pedestrians communicate". In: *2017 IEEE Intelligent Vehicles Symposium (IV).* IEEE. 2017, pp. 264–269.

[144] Ariane Tom and Marie Axelle Granié. "Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads". In: *Accident Analysis and Prevention* 43.5 (2011), pp. 1794–1801. DOI: 10.1016/j.aap.2011.04.012.

[145] Sebastian Hergeth, Lutz Lorenz, and Josef F Krems. "Prior familiarization with takeover requests affects drivers' takeover performance and automation trust". In: *Human Factors* 59.3 (2017), pp. 457–470.

[146] Rajaram Bhagavathula, Brian Williams, Justin Owens, and Ronald Gibbons. "The Reality of Virtual Reality: A Comparison of Pedestrian Behavior in Real and Virtual Environments". In: *Proc. Human Factors Ergonom. Soc. Annu. Meeting* 62.1 (2018), pp. 2056–2060. ISSN: 1541-9312. DOI: 10.1177/1541931218621464.

[147] Kareem Othman. "Public acceptance and perception of autonomous vehicles: A comprehensive review". In: *AI and Ethics* (2021), pp. 1–33.

[148] Lynn M Hulse, Hui Xie, and Edwin R Galea. "Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age". In: *Safety science* 102 (2018), pp. 1–13.

[149] Mayank Bansal, Alex Krizhevsky, and Abhijit Ogale. "Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst". In: *arXiv preprint arXiv:1812.03079* (2018).

[150] Andrea Gorrini, Giuseppe Vizzari, and Stefania Bandini. "Towards Modelling Pedestrian-Vehicle Interactions: Empirical Study on Urban Unsignalized Intersection". In: *arXiv preprint arXiv:1610.07892* (2016). arXiv: 1610.07892.

[151] Michael W Hofbaur and Brian C Williams. "Hybrid estimation of complex systems". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34.5 (2004), pp. 2178–2191.

[152] Xiaoxiao Du, Ram Vasudevan, and Matthew Johnson-Roberson. "Bio-LSTM: A biomechanically inspired recurrent neural network for 3-d pedestrian pose and gait prediction". In: *IEEE Robotics and Automation Letters* 4.2 (2019), pp. 1501–1508.

[153] Christoph Schöller, Vincent Aravantinos, Florian Lay, and Alois Knoll. "What the constant velocity model can teach us about pedestrian motion prediction". In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 1696–1703.

[154] John C. Platt. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods". In: *Advances Large Margin Classifiers* 10.3 (1999), pp. 61–74.

[155] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. "Understanding Pedestrian Behavior in Complex Traffic Scenes". In: *IEEE Transactions on Intelligent Vehicles* 3.1 (2018), pp. 61–70. DOI: 10.1109/TIV.2017.2788193.

[156] Eleonora Papadimitriou, George Yannis, and John Golias. "A critical assessment of pedestrian behaviour models". In: *Transportation Research Part F: Traffic Psychology Behaviour* 12.3 (2009), pp. 242–255. DOI: 10.1016/j.trf.2008.12.004.

[157] Pieter Abbeel, Adam Coates, Michael Montemerlo, Ng Andrew Y., and Sebastian Thrun. "Discriminative Training of Kalman Filters". In: *Robotics: Science and systems*. 2005.

[158] Richard L Knoblauch, Martin T Pietrucha, and Marsha Nitzburg. "Field studies of pedestrian walking speed and start-up time". In: *Transportation Research Record* 1538.1 (1996), pp. 27–38.

[159] Tarak Gandhi and Mohan Manubhai Trivedi. "Pedestrian protection systems: Issues, survey, and challenges". In: *IEEE Transactions on Intelligent Transportation Systems* 8.3 (2007), pp. 413–430.

[160] Yin-Wen Chang and Chih-Jen Lin. "Feature ranking using linear SVM". In: *Causation and Prediction Challenge*. 2008, pp. 53–64.

[161] Nicolas Guuguen, Sebastien Meineri, and Chloe Eyssartier. "A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing". In: *Safety Sci.* 75 (2015), pp. 87–89. ISSN: 18791042. DOI: 10.1016/j.ssci.2015.01.018.

[162] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, and Benjamin Caine. "Scalability in perception for autonomous driving: Waymo open dataset". In: *arXiv* (2019), arXiv–1912.

[163] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John K Tsotsos. "Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 6262–6271.

[164] Boris Ivanovic, Amine Elhafsi, Guy Rosman, Adrien Gaidon, and Marco Pavone. "MATS: An Interpretable Trajectory Forecasting Representation for Planning and Control". In: *arXiv preprint arXiv:2009.07517* (2020).

[165] Xiaobei Jiang, Wuhong Wang, Klaus Bengler, and Weiwei Guo. "Analyses of pedestrian behavior on mid-block unsignalized crosswalk comparing Chinese and German cases". In: *Advances in mechanical engineering* 7.11 (2015), p. 1687814015610468.

[166] Solomon Kullback. *Information theory & statistics*. Courier Corporation, 1997.

[167] Jeff Hecht. "Lidar for self-driving cars". In: *Optics and Photonics News* 29.1 (2018), pp. 26–33.

[168] Daniela A Ridel, Nachiket Deo, Denis Wolf, and Mohan Trivedi. "Understanding pedestrian-vehicle interactions with vehicle mounted vision: An LSTM model and empirical analysis". In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2019, pp. 913–918.

[169] Bastian Jonathan Schroeder. "A behavior-based methodology for evaluating pedestrian-vehicle interaction at crosswalks". PhD thesis. North Carolina State University, 2008.

[170] Somil Bansal, Andrea Bajcsy, Ellis Ratner, Anca D Dragan, and Claire J Tomlin. "A Hamilton-Jacobi reachability-based framework for predicting and analyzing human motion for safe planning". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020, pp. 7149–7155.

[171] W Andrew Harrell. "Factors influencing pedestrian cautiousness in crossing streets". In: *The Journal of Social Psychology* 131.3 (1991), pp. 367–372.

[172] Tianjiao Wang, Jianping Wu, Pengjun Zheng, and Mike McDonald. "Study of pedestrians' gap acceptance behavior when they jaywalk outside crossing facilities". In: *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE. 2010, pp. 1295–1300.

[173] Irtiza Hasan, Francesco Setti, Theodore Tsesmelis, Alessio Del Bue, Marco Cristani, and Fabio Galasso. "" seeing is believing": Pedestrian trajectory forecasting using visual frustum of attention". In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2018, pp. 1178–1185.

[174] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. "Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 14424–14432.

[175] Nishant Nikhil and Brendan Tran Morris. "Convolutional neural network for trajectory prediction". In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018.

[176] Tiffany Yau, Saber Malekmohammadi, Amir Rasouli, Peter Lakner, Mohsen Rohani, and Jun Luo. "Graph-SIM: A Graph-based Spatiotemporal Interaction Modelling for Pedestrian Action Prediction". In: *arXiv preprint arXiv:2012.02148* (2020).

[177] Quinn McNemar. "Note on the sampling error of the difference between correlated proportions or percentages". In: *Psychometrika* 12.2 (1947), pp. 153–157.

[178] Christoph G. Keller, Christoph Hermes, and Dariu M. Gavrila. "Will the pedestrian cross? Probabilistic path prediction based on learned motion features". In: *Joint Pattern Recognition Symp.* Springer Berlin Heidelberg, 2011, pp. 386–395. ISBN: 9783642231223. DOI: 10.1007/978-3-642-23123-0_39.

[179] Mohsen Ahmadi Mousavi, Zainabolhoda Heshmati, and Behzad Moshiri. "LTV-MPC based path planning of an autonomous vehicle via convex optimization". In: *21st Iranian Conf. Elec. Engg.* IEEE. 2013, pp. 1–7.

[180] Jason Kong, Mark Pfeiffer, Georg Schildbach, and Francesco Borrelli. "Kinematic and dynamic vehicle models for autonomous driving control design". In: *IEEE Intell. Veh. Symp.* IEEE. 2015, pp. 1094–1099.

[181] Rico Krueger, Taha H Rashidi, and John M Rose. "Preferences for shared autonomous vehicles". In: *Transportation research part C: emerging technologies* 69 (2016), pp. 343–355.

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

[183] Ivo Batkovic, Mario Zanon, Mohammad Ali, and Paolo Falcone. "Real-time constrained trajectory planning and vehicle control for proactive autonomous driving with road users". In: *2019 18th European Control Conference (ECC)*. IEEE. 2019, pp. 256–262.

[184] Haoyu Bai, Shaojun Cai, Nan Ye, David Hsu, and Wee Sun Lee. "Intention-aware online POMDP planning for autonomous driving in a crowd". In: *IEEE Int. Conf. Robot. Autom.* IEEE, 2015, pp. 454–460. ISBN: 978-1-4799-6923-4. DOI: 10.1109/ICRA.2015.7139219.

[185] S Bertrand, J Marzat, H Piet-Lahanier, A Kahn, and Y Rochefort. "MPC strategies for cooperative guidance of autonomous vehicles". In: (2014).

[186] Orit Bart, Noomi Katz, Patrice L Weiss, and Naomi Josman. "Street crossing by typically developed children in real and virtual environments". In: *OTJR: Occupation, Participation and Health* 28.2 (2008), pp. 89–96.

[187] Sean Vaskov, Hannah Larson, Shreyas Kousik, Matthew Johnson-Roberson, and Ram Vasudevan. "Not-at-Fault Driving in Traffic: A Reachability-Based Approach". In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE. 2019, pp. 2785–2790.

[188] Kihwan Kim, Dongryeol Lee, and Irfan Essa. "Gaussian process regression flow for analysis of motion trajectories". In: *2011 International Conference on Computer Vision*. IEEE. 2011, pp. 1164–1171.

[189] Sanjit A Seshia, Dorsa Sadigh, and S Shankar Sastry. "Formal methods for semi-autonomous driving". In: *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*. IEEE. 2015, pp. 1–5.

[190] Hadas Kress-Gazit, Kerstin Eder, Guy Hoffman, Henny Admoni, Brenna Argall, Ruediger Ehlers, Christoffer Heckman, Nils Jansen, Ross Knepper, Jan Křetínský, et al. "Formalizing and Guaranteeing* Human-Robot Interaction". In: *arXiv preprint arXiv:2006.16732* (2020).

[191] Ryan Felton. *Google's Self-Driving Cars Have Trouble With Basic Driving Tasks: Report*. Aug. 2018. URL: https://jalopnik.com/googles-self-driving-cars-have-trouble-with-basic-drivi-1828653280.