Pedestrian Models for Autonomous Driving
Part II: High-Level Models of Human Behaviour

Fanta Camara[1,2], Nicola Bellotto[2], Serhan Cosar[3], Florian Weber[4], Dimitris Nathanael[5], Matthias Althoff[6],
Jingyuan Wu[7], Johannes Ruenz[7], André Dietrich[8], Gustav Markkula[1], Anna Schieben[9], Fabio Tango[10],
Natasha Merat[1] and Charles Fox[1,2,11]

Abstract—Autonomous vehicles (AVs) must share space with pedestrians, both in carriageway cases such as cars at pedestrian crossings and off-carriageway cases such as delivery vehicles navigating through crowds on pedestrianized high-streets. Unlike static and kinematic obstacles, pedestrians are active agents with complex, interactive motions. Planning AV actions in the presence of pedestrians thus requires modelling of their probable future behaviour as well as detecting and tracking them. This narrative review article is Part II of a pair, together surveying the current technology stack involved in this process, organising recent research into a hierarchical taxonomy ranging from low-level image detection to high-level psychological models. This self-contained Part II covers the higher levels of this stack, consisting of models of pedestrian behaviour, from prediction of individual pedestrians' likely destinations and paths, to game-theoretic models of interactions between pedestrians and autonomous vehicles. This survey clearly shows that, although there are good models for optimal walking behaviour, high-level psychological and social modelling of pedestrian behaviour still remains an open research question that requires many conceptual issues to be clarified. Early work has been done on descriptive and qualitative models of behaviour, but much work is still needed to translate them into quantitative algorithms for practical AV control.

Index Terms—Review, survey, pedestrians, autonomous vehicles, sensing, detection, tracking, trajectory prediction, pedestrian interaction, microscopic and macroscopic behaviour models, game-theoretic models, signalling models, eHMI, datasets.

I. INTRODUCTION

To operate successfully in the presence of pedestrians, autonomous vehicles require input from a huge variety of models that have to work seamlessly together. These models range from simple visual models for detection of pedestrians, to predicting their future movements using psychological and sociological methods. Part I of this two-part survey [33] covered models for sensing, detection, recognition, and tracking of pedestrians. Part II here reviews models for pedestrian trajectory prediction, interaction of pedestrians, and behavioral
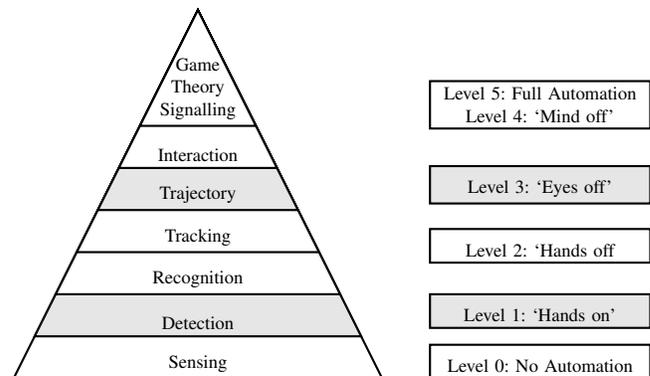

Fig. 1. Main structure of the review

modeling of pedestrians, and also experimental resources to validate all the types of models. Interacting with pedestrians is a particular type of social intelligence. Autonomous vehicles will need to utilize many different levels of models of pedestrians, each addressing different aspects of perception and action. Each of these models can be based on empirical science results or obtained via machine learning. In contrast to the models of Part I, Part II requires models from higher levels of the technology stack, as researched by psychologists and taught in advanced driver training programmes. For instance, drivers often try to infer the personality of other humans, predict their likely behaviours, and interact with them to communicate mutual intentions [100]. Between the high level surveyed in this Part II and the low levels of Part I, researchers infer psychological information from perceptual information. As an example, researchers build systems to recognize the body language, gestures, and demographics information of pedestrians to better predict their likely goals and behaviours. Despite the importance of bridging the research between the higher and lower levels, their connection is still thin, both conceptually and in terms of actual implementations.

While prediction of likely future pedestrian trajectories is becoming increasingly well understood, models for actively controlling pedestrian interactions – including game-theoretic models – are still in their infancy. Active control here means that the vehicle's own future actions are taken into account in predicting how the pedestrian will respond, and vice versa. One reason is that sufficient data to rigorously study interaction between pedestrians has only recently become available as presented in Sec. V on experimental resources. Another

TABLE I
PROPOSED MAPPING FROM SAE LEVELS TO PEDESTRIAN MODEL REQUIREMENTS

SAE LEVEL	DESCRIPTION	PEDESTRIAN MODEL REQUIREMENTS	LOCATION
Level 0 No Automation	Automated system issues warnings and may momentarily intervene, but has no sustained vehicle control.	Sensing	Part I [33] Sec. II
Level 1 Hands on	The driver and the automated system share control of the vehicle. For example, adaptive cruise control (ACC), where the driver controls steering and the automated system controls speed. The driver must be ready to resume full control when needed.	+Detection	Part I [33] Sec. III
Level 2 Hands off	The automated system takes full control of the vehicle (steering and speed). The driver must monitor and be prepared to intervene immediately. Occasional contact between hand and wheel is often mandatory to confirm that the driver is ready to intervene.	+Recognition +Tracking	Part I [33] Sec. IV Part I [33] Sec. V
Level 3 Eyes off	Driver can safely turn attention away from the driving tasks, e.g. use a phone or watch a movie. Vehicle will handle situations that call for an immediate response, like emergency braking. The driver must still be prepared to intervene within some limited time.	+Unobstructed walking models with known goals +Behaviour prediction with known goals +Behaviour prediction with unknown goals	Sec. II-A Sec. II-B Sec. II-C
Level 4 Mind off	No driver attention is required for safety, except in limited spatial areas (geofenced) or under special circumstances, like traffic jams.Outside of these areas or circumstances, the vehicle must be able to safely abort or transfer control to the human.	+Event/Activity Models +Effects of Pedestrian Class on Trajectory +Pedestrian Interaction Models +Game Theoretic and Signalling models	Sec. II-D Sec. II-E Sec. III Sec. IV
Level 5 Full automation	No human intervention is required at all, fully automated driving.	+Extreme robustness and reliability of the models	

Note: '+X' means that model 'X' is required in addition to the models required for the previous SAE level, e.g., SAE level 1 requires 'sensing + detection', SAE level 2 requires 'sensing + detection + tracking'', etc.

reason is that one first has to be able to reliably sense, detect, recognize, and track pedestrians in order to gather enough data for modeling interaction and game-theoretic models. A third reason is that interaction and game-theoretic models are only relevant in crowded environments, while many situations do not require much interaction. However, crowded environments are those that are typically most relevant for autonomous driving. Fig. 1 shows the review structure.

To assess the maturity of the methods presented, the level of autonomy is used, as defined by the Society of Automotive Engineers (SAE) – the same measure has already been used in Part I [33]. For the convenience of the reader, the five SAE levels are briefly presented, ranging from simple driver assistance tools to full self-driving [101]. Requirements for pedestrian modelling increase with each level, with lower levels typically requiring lower and more mature levels of pedestrian models, such as detection and tracking, while higher levels require models for psychological and social understanding to fully interact with pedestrians in a human-like way [30]. Table I gives an overview of SAE levels and requirements mappings.

While many papers propose pedestrian models at various levels, no unifying theory has yet been produced which would make it possible to easily transfer results across all levels from detection to prediction. This review uncovers bottlenecks in transferring results to facilitate closing existing research gaps. Also, many existing studies only consider results from empirical science or those obtained via machine learning. This survey provides an overview considering both possibilities. While machine learning results work particularly well for

detection and recognition, they are not yet performing so well for prediction. Some reasons are that prediction is a more high-dimensional problem, with dimensions including goals, obstacles, various state variables of pedestrians, and road geometry. A further reason is that less labelled data is available for training prediction models. A promising future direction is to combine empirical science results with machine learning to better safeguard techniques using machine learning and to avoid over-fitting.

While similar concepts apply to modelling human drivers and their vehicles for interactions with AVs, this article presents a review of the state of the art specifically in modelling human pedestrians for social decision-making. In some cases it goes beyond modelling aspects to also cover more conceptual aspects or empirical psychological findings, when the studies in question are judged to have very direct applicability to mathematical models. Results from human driving cannot be directly translated to pedestrians due to the variability in locomotion, the differences in shape, the changes in postures and the less-structured environment.

Pedestrians are defined as humans moving on and near public highways including roads and pedestrianised areas, who walk using their own locomotive power. This excludes, for example, humans moving on cycles, wheelchairs and other mobility devices, skates and skateboards, or those transported by other humans. This review does not cover interactions of traffic participants without pedestrians: a survey on trajectory prediction of on-road vehicles is provided in [122] and a survey on vision-based trajectory learning is provided in [145].
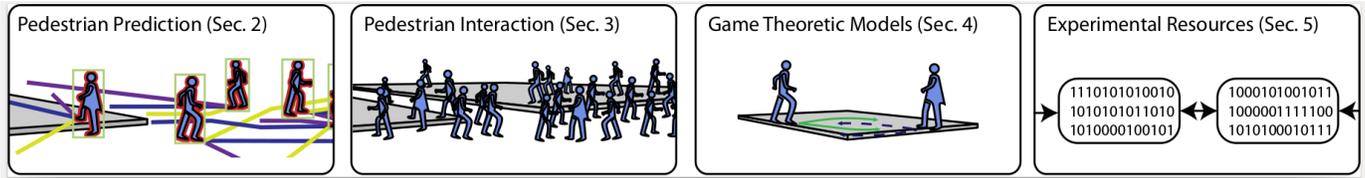
Fig. 2. Structure of the paper

This Part II is organized as shown in Fig. 2. In Sec. II, methods for predicting the movements of pedestrians are reviewed. In particular, we consider models and methods for unstructured environments, for prediction around obstacles, to estimate destinations, and for the prediction of events such as crossing the road. These methods are enhanced in Sec. III for groups of pedestrians interacting with each other. This section considers the complete variety of researched models from macroscopic models only considering flow of people to microscopic models that consider individual pedestrians. In many situations, interaction models do not require game theory, because pedestrians often have different goals. However, there are also many situations, where pedestrians have competing goals, e.g., when several pedestrians have to pass a narrow passage. In such situations, the game theoretic models presented in Sec. IV can be very useful. Finally, Sec. V surveys available resources: datasets and simulators, both for pedestrians and vehicles.

II. BEHAVIOUR MODELS WITHOUT INTERACTION

The tracking models reviewed in Part I are *kinematic* in that they assume that pedestrians move in physical and/or pose space in motion described by kinematic models. This is a very basic assumption – human drivers typically have much more complex understandings and hence predictions of pedestrian behavior which they use to drive safely in their presence [100]. These range from slightly more advanced kinematic understandings such as 'pedestrians tend to walk in straight lines to models of how they are likely to interact with static objects in their environment, and predictions of pedestrians' likely destinations from reading the street scene.

This section reviews such models starting from simple unobstructed path models to uncertain destination models and more advanced event/activity models. These models do not yet consider interaction with other agents. Figure 3 summarizes the classes of models presented in this section. A previous review was proposed by Ridel *et al.* [171], which mainly considered pedestrian crossing intent and offered a restricted view of the different models developed for trajectory prediction.

A. Unobstructed walking models with known goals

Given a start location and orientation, and a goal location, humans do not typically turn towards the goal on the spot (which would waste time) and then walk in a straight line, but rather set off walking in their initial heading and adjust their orientation gradually as they walk, resulting in smooth, curved trajectories from origin to destination [95], [69]. Models from optimal control theory – such as the one proposed in
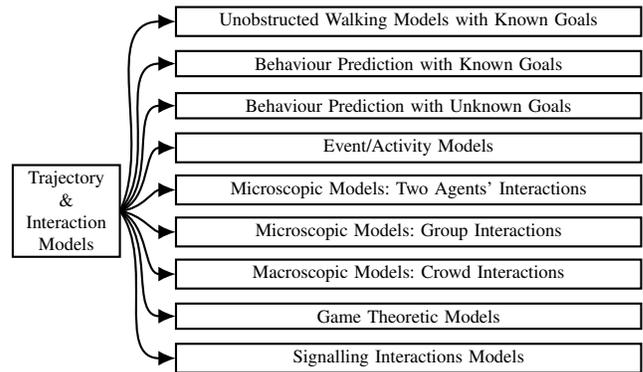

Fig. 3. Pedestrian behaviour prediction and interaction models

[95] as also used in robotics [48] – define cost functions for travel time, speed, and accelerations, to reproduce these characteristic curved trajectories. The model in [69] instead achieves curved trajectories by modelling the rate of turning of the pedestrian as a function of the visual angle and distance to the goal. A simple kinematic model consists in considering human locomotion as a nonholonomic motion [160], using the unicycle model (1) where the pedestrian walking trajectory is represented by the trajectory of their center of gravity, 2D coordinates (x, y) and by the angle θ,

$$\begin{aligned} \dot{x} &= u_1 \cos \theta \\ \dot{y} &= u_1 \sin \theta \\ \dot{\theta} &= u_2 \end{aligned} \tag{1}$$

where u_1 is the forward velocity and u_2 is the angular velocity. Assuming known origin and destination with inverse optimal control, one can reliably predict human walking paths using this model [9] [154].

B. Behaviour prediction with known goals

Here, the likely behaviour of a pedestrian in a static environment is considered, given a map. Pedestrians are likely to route around obstacles, and to stop at the edges of roads before crossing. This section does not consider social effects of other agents – this is presented later in Sec. III.

1) Dynamic graphical models: Dynamic Graphical Models (DGM) are Graphical Models of a particular topology, containing some Markovian sequence of variables over time. DGMs include simple Markov and Hidden Markov Models and also more complex models. The method in [144] used tracking in a DGM based on particle filter approximation to infer beliefs over future pedestrian trajectories and combined this with a GNSS (Global Navigation Satellite System) module that provides information about the hazardous areas and people.

2) Gaussian Process methods: Habibi *et al.* [85] proposed a context-based approach to pedestrian trajectory prediction using Gaussian Processes [165]. This model incorporates context features such as the pedestrian's distance to the traffic light, the distance to the curbside, and the curbside orientation in the transition learning phase to improve the prediction. A context based augmented semi non-negative sparse coding (CASNSC) algorithm is used to predict pedestrian trajectories.

3) Deep learning methods: Bock *et al.* [24] developed a Recurrent Neural Network (LSTM) model to learn pedestrian behaviour patterns at intelligent intersections using camera data from the onboard vehicle and the infrastructure. The model can predict trajectories for a horizon of 5s.

4) Other methods: Kruse *et al.* [120] was one of the first attempts to statistically infer human motion patterns from data and incorporate them in a robot motion planner for obstacle avoidance. Garzón *et al.* [74] presented a pedestrian trajectory prediction model based on two path planning algorithms that require a set of possible goals, a map and the initial position. It then computes similarities between the obtained and observed trajectories into probabilities. This model is run along with a pedestrian detector and tracker. Tamura *et al.* [196] proposed a pedestrian behaviour model that is based on social forces and takes into account the intention of the pedestrian in the trajectory prediction by defining a set of subgoals. In [169] the uncertain goals are used as latent variables to guide the motion prediction of pedestrians. Their positions are predicted by combining forward propagation of a physical model with local a priori information (e.g., obstacles and different road types) from the start position, and by planning the trajectory from a goal position. The distribution over the destinations is modeled with a particle filter.

In [207], Vasishta *et al.* presented a model based on the principle of natural vision that incorporates contextual information extracted from the environment to the pedestrian behavior and it especially tries to predict hazardous behavior such as crossing in non authorized areas. The aforementioned model in [69] considers goals and obstacles as distance-dependent attractors and repellers in heading angle space. The contributions from the goal and obstacles are linearly combined, yielding a momentary rate of acceleration of heading, which results in human-like trajectories for simultaneous goal-seeking and obstacle avoidance. In [55], Dias *et al.* developed a model simulating pedestrian behaviour around corners, using minimum jerk theory and one-thirds power law concept. Their model uses Monte Carlo simulation to generate pedestrian trajectories with turning maneuvers, which were comparable to empirical trajectories.

C. Behaviour prediction with unknown goals

Many of the above models assume known probable destinations for pedestrians, which enable routing to act not just around local obstacles, but to predict entire long-term trajectories, such as for pedestrians intending to cross the road. However, in reality a pedestrian's destination is rarely given.

1) Dynamic graphical models: Ziebart *et al.* [231] presented a pedestrian trajectory prediction model that takes into account hindrance due to robot motion, as is required in off-carriageway interactions such as last mile AVs in pedestrianized areas. A maximum entropy inverse optimal control technique, introduced in [230], is used and is equivalent to a soft-maximum version of Markov decision process (MDP) that accounts for decision uncertainty into the trajectories distribution. The cost function is a linear combination of the features (e.g obstacles) in the environment. People's motion can be modeled by an MDP and by choosing a certain path, there is an immediate reward. The model is conditioned on a known destination location but the model reasons about all possible destinations and the real destination is not known at the prediction time. The destination is inferred in a Bayesian way, by computing the prior distributions over destinations using previous observed trajectories. When there is no previous data, features (door, chair etc.) in the environment are used to model the destination. In [112], Kitani *et al.* extended [230], [231] by incorporating visual features to forecast future activities and destinations. The observations provided by the vision system (e.g. tracking algorithm) are assumed to be noisy and uncertain therefore they used a hidden variable Markov decision process (hMDP) where the agent knows its own states, action and reward but observes only noisy measurements. Negative Log-Loss (NLL) is used as a probabilistic metric and Modified Hausdorff Distance (MHD) as a physical measure of the distance between two trajectories. Vasquez [208] extends the work of Ziebart [231] and Kitani [112] while reducing computational costs.

Bennewitz *et al.* [18], [17] proposed a learning method for human motion recognition using the expectation maximization (EM) and a hidden Markov model (HMM) for clustering and predicting human trajectories and incorporating them into a robot path planner. In [219], Wu *et al.* presented a model that uses Markov chains for pedestrian motion prediction (able to deal with non-Gaussian distribution and several constraints). A heuristic method is proposed to automatically infer the positions of several potential goals on a generic semantic map. It also incorporates policies to predict the pedestrian motion direction and takes into account other traffic participants by incorporating a collision checking approach. Borgers *et al.* [29] presented a model that predicts pedestrians' route choice based on Markov chains. Similarly, Bai *et al.* [11] presented a real-time approximate POMDP (Partially Observable Markov Decision Process) controller, DESPOT, for use in high-street type environments. The method is intention-aware in the sense of inferring pedestrian destinations and route plans from their observed motion over time, and accounting for the value of this information against the value of making progress while planning a robot's own route around them. Karasev *et al.* [109] presented a long-term prediction model that incorporates environmental constraints with the intent modeled by a policy in a MDP framework. The pedestrian state is estimated using a Rao-Blackwellized filter and pedestrian intent by planning according to a stochastic policy. This model assumes that pedestrians behave rationally.

2) Deep learning methods: Hug *et al.* [96] proposed a LSTM-MDL model combined with a particle filter method for multi-modal trajectory prediction, and tested on Stanford

Drone Dataset (SDD) [175]. Rehder *et al.* [170] proposed a method to infer pedestrian destinations. The trajectory prediction is computed as a goal-oriented motion planning. The whole system is based on deep-learning and trained via inverse reinforcement learning. A general introduction on reinforcement learning in robotics can be found in [114]. Deo *et al.* [54] presented a framework for multi-modal pedestrian trajectory forecasting in structured environments. They used a convolutional neural network to compute both the reward maps of the path states and the possible goal states for MDPs. The derived policy information is then fed into a recurrent neural network, combined with track history, to generate possible future trajectories. Goldhammer *et al.* [78] developed a Multilayer Perceptron (MLP) neural network with polynomial least square approximation to predict pedestrian trajectories based on camera data. A long-term prediction model using RNNs is proposed in [22].

3) Other methods: Cosgun *et al.* [50] presented a person-following service robot with a task dependent motion planner. The robot can track and predict the future trajectory of the person by maximizing its reward at future steps while avoiding entering into the human's personal space. Koschi *et al.* [116] proposed a set-based method to predict all possible behaviours of pedestrians using reachability analysis [5] for pedestrian occupancy. Pedestrians are described as point mass with a certain maximum velocity and maximum acceleration. A rule-based occupancy is applied that does not allow a pedestrian to obstruct traffic, e.g. pedestrians are given priority at crosswalks and their trajectory is assumed to be evasive.

D. Event/activity models

Pedestrian event models consider stereotypical sequences of behaviours of individual pedestrians. These may give additional information about route choice, beyond that available from static classification of the pedestrian. For example, a commuter, or class of commuters, who engage in similar actions every day, such as road crossing in a certain way then checking their phone, may reveal information about their identity to enable re-identification[1] which is in turn predictive of their future destinations. These models look for features predictive of route choice in static environments and do not consider social factors.

1) Dynamic graphical methods: Duckworth *et al.* [63] [62] developed on a mobile robot an unsupervised qualitative spatio-temporal relations (QSR) model to learn motion patterns using a graph representation and is able to predict people's future behaviour. Dondrup *et al.* [61] presented a ROS-based real-time human perception framework for mobile robots using laser and RGB-D data and tracking people with a Kalman filter approach. Human trajectories are converted into QSR (Qualitative Spatial Relations) and used for a Hidden Markov Model (HMM) to classify the behaviour of the different people encountered [60]. In [185], Schneider and Gavrila presented a comparative study on Bayesian filters (EKF and IMM) for short-term (<2s) pedestrian trajectory

prediction, in particular they used stereo camera images to apply these methods to four different types of behaviour: crossing, stopping, bending in and starting.

Body heading is used above in basic path planning models, but head-turning events are distinct from body heading, and are discrete events which occur when a pedestrian turns their head to look around rather than to orient their body. Such an activity model is used in [115] to enhance path prediction of pedestrians while intending to cross a street. For low-level occupancy prediction, a dynamic Bayesian network (DFBN) is used on top of a switching linear dynamic system (SLDS) anticipating the changes of pedestrian dynamics. As in [115], studies [187], [188] also model head orientation by an event/activity model to enhance the underlying prediction approach.

2) Gaussian Process methods: Quintero *et al.* [161] [162] proposed a pedestrian path prediction method up to 1s ahead based on balanced Gaussian Process dynamical models (B-GPDMs) and naïve Bayes classifiers. GPDM is used to transform a sequence of timed feature vectors into a low dimensional latent space and it can predict the next position based on the current one. The naïve Bayes classifiers are used to classify pedestrian actions based on 3D joint positions.

3) Feature selection methods: Bonnin *et al.* [27] proposed a generic context-based model to predict pedestrians behavior according to features describing their local urban environment. To learn about interactions between autonomous vehicles and pedestrian interactions, in [37], Camara *et al.* collected data from real-world pedestrian-vehicle interactions at an unsignalized intersection. The actions of pedestrians and vehicles were ordered into sequences of events comprising descriptive features and the study revealed the most predictive features in a crossing scenario such as the head direction, the position on the pavement, hand gestures etc. In [36], these features were filtered over time to predict whether the pedestrian would first cross the intersection or not. Völz *et al.* [211] [212] proposed a model that can predict whether or not a pedestrian will cross the street with a set of features learnt from a database of LIDAR pedestrian trajectories that are used as inputs for a support vector machine (SVM).

E. Effects of pedestrian class on trajectory

The models reviewed so far consider all pedestrians to be alike, but human drivers interacting with pedestrians may consider their attributes as members of stereotypical classes. Membership of various demographic and psychological state classes may be predictive of their behaviour. This section first reviews findings from the psychological literature suggesting what such classes could be usefully predictive of behaviour, if it was possible to classify them automatically from autonomous vehicles. Rasouli and Tsotsos reviewed pedestrian demographics for interactions with autonomous vehicles and argued that knowing such information could help AVs, cf. Sec. III. 1. in [168]. Figure 4 presents a set of pedestrian attributes used for behaviour modelling.

a) Effects of age and gender: Wilson *et al.* [217] performed a large-scale study on adult pedestrian crossing

---

[1]Identity here is distinct from personal information as defined by privacy laws such as the EU General Data Protection Regulation (GDPR).
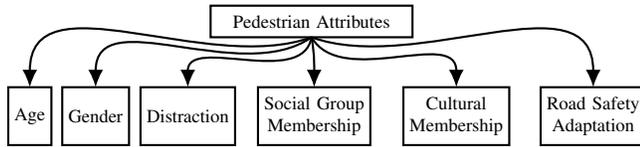
Fig. 4. Pedestrian behaviour attributes

behavior and concluded that elderly people take more time and have more head movements during the crossing. Evans *et al.* [68] used the Theory of Planned Behavior (TPB) [1] via a questionnaire to predict adolescents' intentions during a hazardous road-crossing scenario. Their results show that older and male adolescents had stronger intentions to cross and that moral norms do not have any influence on crossing decisions. Pedestrians who considered themselves as safe pedestrians were less likely to cross and the anticipated affective reactions were important. Bernhoft and Carstensen [21] compared the crossing preferences and behaviour of elderly pedestrians and cyclists (age 70+) to younger people aged 40-49. It was found that elderly people have a preference for road facilities that they consider to be safer such as pavements, pedestrian crossings, signalized intersections, cycle paths. The differences between the two groups are said to be related to health and physical abilities of the people rather than their differences in age and gender.

Several studies have shown that older pedestrians have a larger accident rate than younger people [217]. Gorrini et al. [80] also found differences in adults and elderly people crossing behaviour. The study of Oxley *et al* [152] showed that older pedestrians have more risky crossing behavior in complex traffic environments than younger people. Not surprisingly, many authors have found decreasing crossing speeds with age [10] [129], [200], compensated for by requiring larger time gaps in traffic before commencing crossing [129]. In addition, Avineri *et al.* [10] found lower crossing speeds for female than male pedestrians, and that the fear of falling in elderly pedestrians has an effect on the number of downward head pitches during crossing. Holland and Hill [92] used the TPB for pedestrians' intention analysis while crossing the road. The results showed that women perceived more risk and were less likely to cross than men. In [93], they also studied the effect of gender on pedestrian crossing behaviour and showed that men with a driving experience make safer crossings than non-drivers and that older women were found to make more unsafe crossing decisions than younger women.

b) Distraction: Distraction of pedestrians from traffic environments would ideally be defined via their mental state i.e., thinking about a problem unrelated to their environment; or approximated in practice via observable proxies. While it is possible that mental distraction might be measurable via hard-to-observe proxies such as gaze direction or high-level body language, it may be more practical to look instead for known *causes* of distraction. Schwebel *et al.* [189] performed a study in a semi-immersive virtual pedestrian street with college students, finding an impact of talking on mobile phones on crossing behaviour. Walker *et al.* [213] showed that male

pedestrians using a personal music device were more cautious in crossing than those who were not distracted. In [198], the effects of personal electronic device usage on crossing behavior is studied. The results show a third of the observed pedestrians were distracted by their mobile phone and that distracted pedestrians are more likely to have unsafe crossing behaviour and walk much faster than undistracted pedestrians.

c) Social group membership: Group membership can affect road crossing. Three strangers in a group are less likely to assert in a crossing than three friends. In particular, group size influences a lot crossing behavior [168]. Zeedyk *et al.* [224] performed a study with adult-child pairs while crossing the road at a pedestrian crossing. They found that adults were more likely to hold girls' hands than boys.

d) Cultural membership: In contrast to the above membership of short-term, physically present groups, it is also possible to consider 'cultural membership' of a pedestrian to any long-term, non-physically present group that may be usefully predictive of behaviour. For example, it might be possible for a human driver or autonomous vehicle to classify pedestrians as members of religious, sporting, or musical (sub)cultures as a probabilistic function of features of their clothing such as shape and colour of garments or symbols displayed on them; and that members of such groups show statistically significant differences in assertiveness, politeness, and other road interaction behaviours (cf. [168]). In Sociology, classifications of individuals into cultures is notoriously problematic and politicised. But for the purpose of predicting road interactions, *any* classification derived from observable features may be usefully considered if it improves predictions.

e) Road safety adaptation: Related to the possible predictiveness of cultural clothing is the effect of road safety clothing on behaviour. Human drivers are more likely to yield to pedestrians wearing high-visibility clothing [89], so it is also possible that knowing this fact will make a pedestrian wearing such clothing more likely to behave assertively. This is an example of risk compensation *adaptation*, a well-known effect in road safety in which the owners of safety improvements make economic decisions whether to use them to reduce accidents or alternatively to gain some other advantage at the cost of retaining the original accident rates [179].

F. Discussion

Single pedestrian unobstructed walking path and behaviour prediction around obstacles for known origins and destinations has well-established solutions. Their main strength lies in their simplicity and ease of implementation but their applicability to solve real AV problems is very limited due to the strong assumptions (e.g static obstacles, known origin-destination of pedestrians) which are not easily verified in the real world. But when – as is usual in real-time systems – the destinations of pedestrians are not known in advance, trajectory prediction is harder and remains an open research area.

Uncertain destination models may use known destination models as a subcomponent and average over them weighted by predictions about what the destination is. To predict what a pedestrian's destination will be, many medium and high-level sources of information may be relevant and useful, if

suitable models can be found. These models split roughly into short-term models for prediction horizons around 1-2s and long-term models predicting for a horizon of around 5-6s. Event-based models of activity assume that behaviour often contains repeated stereotypical chunks of behavior, which once recognised in early stages can predict their later stages. The major emerging long-term prediction methods rely on neural network ('deep learning) methods. There is a need to verify how the data-driven methods such as [6] can be actually applied online for real-time systems. These models can help AVs to more accurately predict single pedestrian behaviour for shorter or longer time horizons, e.g. to know precisely whether a pedestrians trajectory would interfere with the AV's own path. But their main challenges lie in their computational cost, which increases significantly with the number of destination guesses, with longer time horizons and the amount of data needed for learning pedestrian motion patterns. Moreover, deep learning models are sometimes referred to as 'black-box models, in the sense that AI developers cannot fully explain some decisions (e.g. feature selection) made by the neural networks, rendering them potentially problematic for investigating the causes of incidents involving AVs and for determining their liabilities [41][81].

Single pedestrians destinations and behaviours may be informed by their class memberships, including their demographics and other visible features, such as clothing types. There are many recent sociological studies giving evidence of these effects, but they have not yet been translated into algorithms suitable for autonomous vehicle use, which would be a promising new research area. It is conjectured that additional information about pedestrians' emotion states would be similarly informative (e.g. angry pedestrians more likely to assert themselves in competitions for road space), but no studies were found in this area. Traditionally, emotional state has been difficult to capture and record, so that manually annotation of data sets are too small for machine learning to use. But as machine vision for face and body language recognition continues to improve (cf. Part I [33] Sect. IV), they are expected to produce big data sets which will enable machine learning to operate and inform destination and behaviour predictions.

III. PEDESTRIAN INTERACTION MODELS

So far, only path prediction models for single pedestrians in static environments ignoring interactions with other pedestrians have been reviewed. This section will consider models of interaction between pedestrians. In Social Science, pedestrian behavior models have been studied for a long time: a survey is provided in [40] [199]. These models can be classified in two categories, namely microscopic models and macroscopic models, as reviewed in [209]. Microscopic models model only each pedestrian individually. Macroscopic models do not model individual pedestrians and instead model the behaviour of a single aggregate entity such as a crowd or a flow. Papadimitriou *et al.* [153] presented a review on pedestrian behavior models and a study on pedestrian and crowd dynamics was proposed by Vizzari and Bandini in [210]. Bellomo
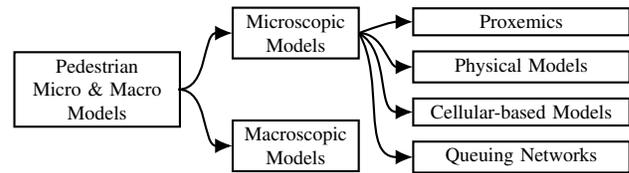

Fig. 5. Pedestrian microscopic and macroscopic models

et al. [15] reviewed mathematical models of vehicular traffic and crowds while Duives *et al.* [64] surveyed pedestrian crowd simulation models. Figure 5 presents a summary of pedestrian microscopic and macroscopic models.

A. Microscopic models

This section first describes pedestrian behaviour models at the microscopic level. It then presents pedestrian interaction models using these behaviour models for two agents interactions and group behaviour modelling.

1) Behaviour models: Microscopic models are divided into three main groups: physical models, cellular-based models and queuing network models. Each model is generally structured by two terms: one term that represents the attractive effects of pedestrians toward their goal and the other repulsive effects among and between pedestrians and the obstacles [40]. Proxemics is first described in this section.

a) Proxemics: The Psychology theory of Proxemics [88] studies human preferences (utilities) for having other humans in their proximity. Proxemics typically identifies four radial comfort zones, whose radii differ between cultures, for intimate, personal, social, and public space. These zones can be described by eight dimensions [88]:

1) postural-sex identifiers
2) sociofugal-sociopetal orientation (SFP axis)
3) kinesthetic factors
4) touch code
5) retinal combinations
6) thermal code
7) olfaction code
8) voice loudness scale

This model has been empirically tested with participants [215]. The theory is of great interest to pedestrian interaction models because it provides a possibly hard-wired negative utility not just for actual collisions with pedestrians but also for simply feeling too close to them. In particular, this provides a method for an AV to inflict a real negative utility on a pedestrian without touching them or risking their physical harm. Binary proxemics is the simplest case used in simple models, in which a negative utility is assigned to actually hitting someone, and zero utility is assigned to not hitting anyone. Zonal proxemics is more subtle, it relies on the eight proxemic dimensions defined above. It assigns different utilities to the presence of a person in four different zones around an individual which are defined as the *intimate distance*, the *personal distance*, *social distance* and the *public distance* [87]. Gorrini *et al.* [79] studied the proxemics behaviour of groups of pedestrians in interaction and showed that it has negative

effects in walking speed for evacuation scenarios. Manenti *et al.* [138] presented an agent-based pedestrian behaviour model that takes into account proxemics and group behaviour. Their model was tested with groups of people and in a simulated environment. A detailed review on proxemics models for robot navigation among humans is proposed in [172].

b) Physical models: These are splitted into three sub-categories. The *utility maximization model*, as used in [113], assumes that pedestrian behaves such as to maximize their utility, for example their speed of motion and approach or avoidance of some objects or persons. In the *magnetic force model* proposed by [150], the pedestrian behavior is determined by the equation of motion of the magnetic field. Pedestrians are positive poles and their destinations are negative poles. In the *social force model*, introduced by Helbing [90], each pedestrian has a desired velocity, a target time and a target destination which are affected by social forces such as the interaction with other pedestrians and the effects of the environment. In [132] social forces are described as individual forces (fidelity, constancy) and group forces (attraction, repulsion, coherence). Most of the time, social forces are modeled such that to minimize an energy objective which include terms for individual and group forces.

c) Cellular-based models: These represent a cost model such as Blue and Adler's cellular automata model [23] and used for motion prediction. Cellular Automata (CA) is a discrete, time based modeling formalism on a regular cell grid. It describes the walk of a pedestrian according to rules of a cell occupancy, e.g. a cell can be occupied only if it is free and a pedestrian can have three possible movements: lateral, longitudinal or mitigation of the conflicts. The benefit cost model, developed by Gipps and Marksjo [76], is a discrete and deterministic model where the space is divided into a grid of cells and each agent is described as a particle in a cell. A benefit value, equivalent to the pedestrian utility, is arbitrarily assigned to each cell. In [58] a cellular automata model simulates multi-agent interactions.

d) Queuing network models: They have been developed for studies of evacuation dynamics [133]. These are evaluated by Monte Carlo simulation methods for discrete events. Each pedestrian is represented as an individual flow entity interacting with other objects, facilities are modeled as a network of arches for openings and of nodes for rooms. In [13], a queuing network model is compared to a social force model for pedestrian crossing movement prediction.

2) Two agents' interaction: These models are those involving only two agents with mutually influencing behaviours, rather than larger groups of agents. They may be simpler than larger group models but sometimes provide a foundation for extension to larger group models, hence they are here presented first.

a) Dynamic graphical models: The method in [31] uses POMDPs (Partially Observable Markov Decision Processes) with a time-indexed state space to model interactions and they used the example of an elevator-riding task to test the model. In [178], Rudenko *et al.* proposed a method that uses MDPs with a joint random walk stochastic policy sampling algorithm to predict motion and social forces to model interactions. The

model in [119] learns features from observed pedestrian behaviors using a Markov Chain Monte Carlo (MCMC) sampling and performs a Turing test with human participants to validate the human-like behavior of the model. Chen *et al.* [46] used an extended Kalman filter to predict future motions of pedestrians and estimate the time-to-collision range (TTCR) for collision risk level identification.

b) Gaussian Process methods: Kawamoto *et al.* [111] proposed a method to learn pedestrian dynamics with kriging, the most traditional form of Gaussian Processes. Their work can predict pedestrian movement using spatial kriging and spatio-temporal kriging. Social interaction is modeled by spatio-temporal correlation of pedestrian dynamics and correlation is estimated by kriging.

c) Deep learning methods: Alahi *et al.* [2] predicted pedestrian trajectories in crowded spaces using a 'social LSTM, a variant of recurrent neural network model that can learn human movement (velocity, acceleration, gait...) taking into account social human motion conventions and predict their future trajectories. This technique is opposed to traditional social forces methods and outperforms most the state-of-art methods on public datasets (ETH and UCY). Long Short-Term Memory (LSTM) can learn and reproduce long sequences, it is a data-driven technique. One LSTM is used for each person and the interaction among people is modeled by a social pooling layer which allows the share of states between neighboring LSTMs. Although group behavior is not modeled, the social LSTM can predict it very well. Similarly to the previous method, Chen *et al.* [47] developed a long-term pedestrian prediction model using RNNs for pedestrian trajectory prediction.

d) Road crossing models: This section extends the event-activity models from section II-D by adding interaction between pedestrians and vehicles. When microscopic models of pedestrian movement are included in larger-scale traffic simulations together with vehicles, they are typically extended with specific provisions to account for pedestrian's decisions on where and when to initiate road crossing, when this is needed for the pedestrians to reach their goals. Other, so called *gap acceptance* models, have instead described probabilities of pedestrians crossing in a certain gap between vehicles, using generalised linear models, with predictors including both the available gap itself, as well as other factors such as age and gender of the pedestrians, number of pedestrians waiting to cross, and time spent waiting [193], [186].

Markkula *et al.* [139] proposed another type of model for pedestrians road crossing decision, modelled as the result of a number of perceptual decisions concerning the available gap, but also car yielding, explicit communicative signals from the car, and eye contact with the driver. These decisions were described as several interconnected evidence accumulation processes, and it was shown that empirically observed bimodal distributions of pedestrian waiting time were qualitatively reproduced by the model. In [38], Camara *et al.* proposed a heuristic model for pedestrian crossing intention estimation. Their method is based on a distance ratio model that computes the pedestrian crossing probability over time until the curbside. Their results showed that this heuristic model is sufficient for

most of the crossing scenarios present in the dataset used and that the remaining scenarios would require higher level models such as game theory.

e) Other methods: Discrete choice models in [8] [28] offer a framework to model pedestrian walking along link levels, where pedestrian paths are composed of a sequence of straight lines in absence of any obstacles. For example, the model in [28] has been applied to predict pedestrian behaviour in the presence of other people in shopping street areas. Hoogendoorn and Bovy [94] extended a standard optimal control theory model of a free walking pedestrian, X, to include their first and second order predictions of another pedestrian Y, from the point of view of an external observer, O. It is important to understand the points of view involved here, because this consideration motivates higher order and game theoretic models of behaviour. In the 'single pedestrian models above, the model is made from O's point of view and assumes the pedestrian X does not respond to anyone else's actions in any way. In O's first-order model of X encountering Y, the model includes X's own single-pedestrian model of Y's motion, and X's likely responses to it. For example, X estimates that X's path will collide with Y's kinematic path so alters X's path to avoid Y. In O's second-order model of X's behaviour: X is modelled not only as doing this but also as modelling Y's model of X's kinematic behaviour, predicting Y's likely action based on that, and planning X's own action accordingly. These first and second order terms are added as costs to the standard free walking model. They are however incomplete, because one could easily imagine third and fourth order models and so on, forming an infinite sequence of additional cost terms. However, Hoogendoorn and Bovy found in simulation that the second order model produces flows of pedestrians in crowded environments similar to those observed in some Japanese crossings.

3) Group interaction: A *group* is here considered to be a collection of more than two pedestrians, but smaller and more cohesive than a crowd. These models are developed primarily for use by non-carriageway autonomous vehicles, such as delivery robots, navigating through crowded pedestrianized areas, needing to cut their way between groups.

a) Dynamic graphical models: In [19] a real-time pedestrian path prediction is performed in cluttered environments without making any assumption on pedestrian motion or pedestrian density. Pedestrian motion and movements patterns are learnt from 2D trajectories. Bera *et al.* used sparse and noisy trajectories data from indoor and outdoor crowd videos. By combining local movements (microscopic and macroscopic motion models) and global movements (movement flow), the patterns help improve the accuracy of the long-term prediction. An ensemble Kalman filter (EnKF) was used to predict the next state based on current observation and EM algorithm to maximize the likelihood of the state. Pedestrian clusters are computed based on their positions, velocities, inter-pedestrian-distances, orientations etc. Global movement patterns are the past movement and intended velocity of pedestrians. Local movement patterns are obtained by fitting the best motion model to pedestrian clusters and individual motions. In [20], the same authors implemented a tracking algorithm built on

top of [19]. Deo *et al.* in [53] uses VGMMs to model pedestrian trajectory using pedestrian origins and destinations. Their model is tested on a dataset of a crowded unsignalized intersection in a university campus. Pellegrini *et al.* [155] introduced a linear trajectory avoidance (LTA) model which has similarities with the social force model. In [156], the same authors extended the LTA model with a stochastic version taking into account group behavior and allows multiple hypotheses about the pedestrian position. Zhou *et al.* [229] proposed a mixture model of dynamic pedestrian-agents (MDA) for pedestrian trajectory prediction in crowds.

b) Gaussian Process methods: Henry *et al.* [91] used inverse reinforcement learning (IRL) to learn human-like navigation behavior in crowds. The model estimates environmental features using Gaussian Processes and extends Maximum Entropy Inverse Reinforcement Learning (MaxEnt IRL) of [230] by assuming that features in the environment are partially observable and dynamic. The proposed approach was developed for mobile robot motion planning, but it could be used for human motion prediction. In [201], Trautman and Krause proposed to solve the freezing robot problem, where a robot motion planner gets stuck and cannot find any proper move to perform, by a model based on Gaussian Processes, a statistical model that is able to estimate crowd interaction.

c) Deep learning methods: The subsequent models may not explicitly consider interaction, but they learn interaction implicitly through machine learning techniques. The model in [194] implemented a real-time Temporal 3DOF-Pose Long-Short-Term Memory using 3D lidar data from a mobile robot. Shi *et al.* [191] developed a long-term pedestrian trajectory prediction model for crowded environments using LSTM. In [222], Yi et al. proposed a deep neural network model called behavior-CNN that is trained with crowded scenes video data. A pedestrian behavior model is encoded from the previous frames and used as an input for the CNN model to predict their future walking path and destination as well as a predictor for a tracking system. Radwan *et al.* [163] presented an interaction-aware TCNN, a convolutional neural network model that can predict interactive motion of multiple pedestrians in urban areas.

Amirian *et al.* [6] predicted the motion of pedestrians over a few seconds, given a set of observations of their own past motion and of those of the pedestrians sharing the same space, using a Generative Adversarial Network (GAN)-based trajectory sampler. The reason for this choice is that such a method naturally encompasses the uncertainty and the potential multi-modality of the pedestrian steering decision, which is of great importance when using this predictive distribution as a belief in higher level decision-making processes. Lee *et al.* [121] developed *DESIRE* a trajectory prediction framework for multiple interacting agents based on deep neural networks. A conditional variational auto-encoder is used to generate hypothetical future trajectories. An RNN is then used to score and rank those features in an inverse optimal control manner and taking into account the scene context. Gupta *et al.* [82] proposed a socially-aware GAN with RNNs for pedestrian motion sequence prediction in dynamic environments. However, their model assumes that people influence each

other uniformly. A detailed analysis and improvement of this GAN method is proposed in [117]. With a similar method, called SoPhie, Sadeghian *et al.* [180] developed a GAN-based trajectory prediction model that focuses on the most important agents for each interacting agent.

d) Other methods: Moussaid *et al.* [147] presented a heuristics-based model to predict pedestrian behavior in crowded environments. Based on the idea that visual information is very important for pedestrians [12], [202], they found that two simple heuristics can model the interaction among people: the desired walking direction and speed of pedestrians are sufficient. Bonneaud and Warren [26] proposed a related type of model, extending the behavioral dynamics model by [69] to goal-seeking and obstacle avoidance in crowds, and found that the model was able to reproduce qualitative crowd phenomena like lane formation. The model in [99] learns behavioral patterns from pedestrian trajectories in a mall. It assumes that a robot can model interactions using social forces and segment pedestrian trajectories into sub-goals to estimate their future positions.

B. Macroscopic models

In macroscopic models, the crowd is modeled as a single ontological object, replacing and simplifying the representation of multiple microscopic pedestrians. The crowd behaves as a continuous fluid with a flow average speed [197].

The first macroscopic models of pedestrians are due to Hughes and Henderson [97]. The fluid dynamic model classifies pedestrians into groups which are characterized by average features, their position, speed and intended velocity. In [14], pedestrian flows are modeled in simulations for crowded environments. Crowd modelling has also an established community focused on models for evacuation, as reviewed in [182]. In [4] Ali *et al.* used Lagrangian Particle Dynamics to segment high density crowd flows. This method, based on Lagrangian Coherent Structures (LCS) from fluid dynamics and particle advection, is capable of detecting instabilities in the crowd.

Smooth Particle Hydrodynamics (SPH) is a hybrid of microscopic and macroscopic models. Pedestrians are considered individually, but at each time they are aggregated into a density where each particle is moved according to the macroscopic velocity. Etikyala *et al.* [67] reviewed smooth particle hydrodynamics pedestrian flow models while [223] proposed a generic SPH framework for modeling pedestrian flow.

C. Discussion

The theory of proxemics has been well studied in psychology and now being more and more used for VR experiments [151] [56] and computer scientists are just beginning to apply it to make more detailed models of the utility of pedestrian's personal space than simply collisions and non-collisions. In general, microscopic models are preferred to macroscopic models, in particular the social force model is very popular for pedestrian interaction modelling, while macroscopic models are more suited for crowd behaviour modeling, especially in the specialised domain of emergency evacuation modeling. Physical models bring interesting results when there are a lot of interactions, e.g. modelling pedestrian movement in cities [176]. Cellular-based models are useful for modelling pedestrians with minimal movement choices and when representing their collisions is not required. Two agents' and group interaction models offer more precise pedestrian models but they require more computational resources, in particular dynamic graphical, Gaussian Process and deep learning models. More computational research is needed in interaction modelling: psychology/human factors studies and theories are more mature, but their results have not yet been quantified to the extent of enabling translation into algorithms for AVs.

IV. GAME THEORETIC AND SIGNALLING MODELS

A. Game theory interaction models

The interaction models above are all incomplete in the sense that they are of finite order. In the first-order, X's beliefs about Y's future motion are modelled. A second-order model extends to model X's beliefs about Y's beliefs about X's future motion, and so on. But this approach leads to an infinite regress, where there is a need for an infinite order of models to consider X's belief about Y's belief about X's belief about Y's belief and so on. Game theory provides an alternative and stronger framework which does not suffer from infinite regress and allows all considerations by and about all players to be taken into account together.

Isaacs [105] introduced vehicle-pedestrian interactions as the famous 'homicidal taxi driver problem which considered the inverse of the modern AV interaction problem: how an AV controller should act in order to *hit* a pedestrian (the taxi scenario was used initially as a declassification technique to publish missile-defence algorithms, which require control of one missile to hit another). Game theory is in common use in descriptive road user modelling as reviewed in [65], where applications include modelling of lane changes and merging onto motorways, route selection and departure time in congested networks, and socio-economic choices such as purchasing large vehicles or using conventions such as headlight dipping. It has been applied to AV-vehicle interactions in [164] though here only pedestrian models are considered.

The use of game theory for active control of AVs is less common. Descriptive models may be incomplete as active controllers, in particular by allowing for multiple Nash Equilibria to exist without selecting between them. A Nash equilibrium is a set of probabilistic strategies to be played by each of the players, such that no player would change their strategy if they knew the strategies of the other players. It is generally agreed in Game Theory that it is not optimal for players to employ strategies which are not Nash equilibria, though there is still philosophical debate over what strategy is optimal when multiple equilibria exist.

1) Two agents' game theory interactions: The methods in [140] and [204] model selection of pedestrian trajectories from a finite set as a single-shot game. For a small set of known origins and destinations, optimal freespace trajectories are computed from control theory, and actual trajectories from a video set are compared to them and assigned costs according to their deviations from them. These models assume that the

choice of the entire continuous trajectory is drawn from a finite set of previously observed and costed trajectories as a single decision at the start of the interaction and does not model responses to the other agent during the interaction. They are used only as descriptive models rather than as real-time control because they require each pedestrian's final goal location to be known in advance to form the cost matrix – which is only obtainable by looking ahead in the data to see what happened *post hoc*. The authors state that (in the context of AV control), 'few researchers have considered interaction between (pedestrian) objects, thus neglecting that humans give way to each other. Turnwald *et al.* [203] adds an alternative model where one player chooses their trajectory first then the second chooses theirs in response to seeing their initial motion.

Ma *et al.* [134] proposed a long-term game-theoretic prediction of interacting pedestrian trajectories from a single starting image. For each future time in the prediction sequence, fictitious play is used to converge the probabilities of each pedestrian's actions to one (of possibly many) Nash equilibrium. The fictitious play assumes that each pedestrian has a known destination goal, some known visual features (age, gender, initial body heading etc) and a known utility function. The utility function scores vectors of word-state features which contain all of (1) the pedestrian's own future trajectory (which may include control theory style costs); (2) probabilistic beliefs about the other agents' trajectories; (3) the pedestrian's own visual features (age, heading etc); (4) proximity to static obstacles; (5) the pedestrian's distance to their goal. Unusually, the utility functions are learned entirely automatically from video data of actualized trajectories, rather than set by theories. Where theory-like behaviours such as proxemics and social forces are observed in simulations, they arise entirely from this learning process. The functions are assumed to be a weighted linear function of the features and a reinforcement-learning-style model is used to obtain per-state values from the full trajectories during learning. A (deep learning) classifier is used to obtain the visual demographic and heading features from annotated training examples. Performance is degraded when the pedestrian's goal locations are not known and are set to be completely uncertain in the feature vectors.

In [72], Fox *et al.* presented a version of the game-theoretic 'game of chicken for autonomous vehicle-pedestrian interactions at unsignalized intersections. The obtained discrete model called the Sequential Chicken model allows two players to choose a set of two speeds: decelerate or continue. A new method to compute Nash equilibria is presented, called 'meta-strategy convergence, used for equilibrium selection. Camara *et al.* [39] evaluated the model [72] by fitting one parameter θ to controlled laboratory experiments where pedestrians were asked to play Sequential Chicken. This behavioural parameter θ is found to be a ratio between the utility of avoiding a collision and the utility of saving time.

2) Small group game theory models: Vascon *et al.* [206] proposed a game theory model for detecting conversational groups of pedestrians from video data, based on the socio-psychological concept of an F-formation and the empirical geometries of these formations. Johora and Müller [108] proposed a three-layer trajectory prediction model composed of a trajectory planner, a force-based (social force) model and a game theoretic decision model. The game theory model is based on Stackelberg games, a sequential leader-follower game where pedestrians have three different possible actions: continue, decelerate and deviate and the car has two possible actions: continue and decelerate. This model is able to handle several interactions at the same time.

3) Crowd game theory models: Mesmer *et al.* [142] modelled pedestrians' decision-making and interactions during evacuations with game theory. In [190] a model of pedestrian behavior in an evacuation used game theory and showed that pedestrians get greater benefits by cooperating.

B. Signalling interaction models

Signalling models extend interaction models by allowing both the pedestrian and the AV to model and predict each other's actions of giving and receiving pure information, rather than communicating only through their physical poses.

Nathanael *et al.* [148] has proposed a stratified model of mutual awareness between pedestrians and vehicles including AVs. The actor's awareness is divided into three levels, i.e., (1) unaware of the others, (2) factually aware of the other, or (3) aware and actively attending to the other. When one of the two agents is unaware of the other, the interaction may be as simple as collision avoidance by the one aware, relying only on bodily and kinematic cues. When both agents are aware of each other, the interaction takes the form of mutual coordination through implicit cues, whereas when both agents are attentive to each other (as evidenced through eye contact between human actors), the interaction may involve direct communication through explicit signals, such as gestures, nodding etc. In addition attentiveness, as opposed to mere awareness, designates that any physical action from an attentive agent is a response explicitly addressed to the agent at the focus of attention (i.e. it also has a signalling function).

This line of research raises an epistemological question about signalling-based interaction. Some of the models above involve the concepts not just of an agent (1) knowing that the other agent is there, and (2) acting to show the other agent that they are present; but also higher-order knowing and showing these facts. This includes (3) knowing that the other knows they are there and (4) showing the other that they know that the other knows they are there. But also includes arbitrarily higher orders, such as 'knowing that the other has showed that they know that they know that the other knows and so on. There appears to be a potentially infinite regress here, though intuitively most humans find it difficult to comprehend many more levels than the four mentioned here. But it is difficult to argue for why any cut-off should occur at this or other specific level. Intuitively: when two agents make eye contact, they assume that they both then come to know the infinite stack of such statements about each other.

1) Signals from pedestrian to vehicle: The need for precise eye contact as opposed to simple head direction or gaze towards the vehicle is controversial. Considering gaze or head orientation towards vehicles, there is evidence that pedestrians who initiate crossings without looking at the oncoming vehicle

tend to make drivers more attentive to them by keeping larger safety margins [110]. On the other hand, eye contact between pedestrian and driver tends to increase the probability of the vehicle yielding for pedestrians [84]. The apparent controversy between these findings may be attributed to profound differences in the function of these two behavioural traits. While head orientation towards vehicles typically signifies pedestrian situational awareness to drivers, eye contact most probably signifies driver awareness of the pedestrian to the latter [167]. In addition, eye contact is reported to play a non-trivial role in the social dynamics between the two. Nathanael et al. [148] in a naturalistic study of driver pedestrian interaction reported that pedestrian head turning towards a vehicle was sufficient for drivers to confidently infer pedestrians intent in 52% of interaction cases observed. In retrospective think-aloud sessions of their interaction with pedestrians, drivers mentioned pedestrian active head movement and orientation as an important indication of pedestrian awareness of their vehicle. Mutual eye contact between driver and pedestrian was observed only in 13% of interaction cases, accompanied by explicit signalling in 2% of total cases. This is consistent with recent research [166] that reported head orientation/gaze towards vehicles as the most prominent cues for predicting pedestrian intent. In addition, computational models have shown that head direction is a useful trait for pedestrian path prediction and state of situation awareness such as in [37] which argued that if a pedestrian looks at the vehicle, they are less likely to cross the road.

Matthews *et al.* [141] studied pedestrians' behavior with an autonomous goal car equipped with an Intent Communication System (ICS) based on Decentralized MDP to model the uncertainty associated with pedestrian's behavior. Another important factor to take into account is the poor pedestrian signal settings. It has been proven that signal indication and timing affect significantly pedestrian behavior and their crossing decisions [3] [102] [103]. Pedestrians can have sudden speed change while crossing, and such sudden behavioral changes may not be expected by conflicting vehicles, which may lead to hazardous situations. In [103], Iryo-Asano and Alhajyaseen proposed a discrete choice model and Monte Carlo simulation for generating pedestrian speed profiles at crosswalks. In [104], the same authors modelled pedestrian behaviour after the onset of pedestrian flashing green (PFG) via a Monte Carlo simulation. Their results showed a higher probability of pedestrian stopping at longer crosswalks and a significant difference in pedestrian speeds.

Some early steps have however been taken towards modelling at least some levels of explicit knowing and showing of beliefs about each other via signalling behaviour.

2) Signals from vehicle to pedestrian: Beyond understanding pedestrians signalling behavior, game theoretic models may also enable the AV to give signals to the pedestrians, creating a higher level information game with both players communicating through both their physical actions and also their signals. The full game theory of such interactions has not yet been worked out, and will form part of a complex socio-technical system [174], but there has been notable activity – especially via company patents – in researching displays and

other mechanisms for the signalling itself.

Lundgren et al. [131] showed that the lack of two-way communication between driver and pedestrian may reduce pedestrians' confidence to cross the street and their perceived feeling of safety, when crossing. Lichtenthäler *et al.* [128] reviewed robot trajectories among humans, including identifying needs for additional gestures or motion information such as gaze to communicate intention, which is relevant for last mile delivery. Researchers are currently conducting studies to better understand exactly which information needs to be transferred when interacting with an AV. Schieben et al. [183] propose the following information to be considered by the design team.

- Information about the vehicle automation status
- Information about next manoeuvres
- Information about perception of environment
- Information about cooperation capabilities

To transfer the relevant information, two means of communications can be used for shaping the communication language of an AV. First, pedestrians might benefit from direct communication through the means of external human machine interfaces (eHMIs) [131], [177]. Secondly, also careful design of vehicle movement can be used to explicitly communicate. Risto et al. introduced the term 'movement gestures and found 'advancing', 'slowing early and 'stopping short as commonly used gestures [174]. Consistent with this, Portouli et al. [159] in the context of driver-driver interaction have shown that 'edging was explicitly used by drivers trying to enter a two-way street as a sign of their intent to inform oncoming cars. Studies of human robot interaction have shown that allowing humans to anticipate robot movements by explicit communication through movements of the robots head raises perceived intelligence of the robot even if it did not succeed completing its intended tasks [195], thus overcoming potential machine error through the means of explicit communication. These studies might suggest similar devices such as head-like and eye-like displays for AVs.

While Clamann et al. [49] found mixed influences of explicit communication through novel eHMI on crossing behavior in dynamic traffic situations and argued that pedestrians will largely rely on legacy behavior and not on eHMIs, Habibovic *et al.* [86] found that traffic participants feel calmer, more in control and safer when an eHMI was present on an AV. Petzoldt, Schleinitz, and Banse [157] found that an eHMI can help to convey the intention of a vehicle to give priority to a pedestrian. They also observed that pedestrians needed more time to understand the intention of a vehicle without eHMI in mixed traffic situations [157]. Communicating the intent and awareness of automated vehicles has been considered in a positive way [136] [137]. Habibovic et al [86], [7] argued that, for safety reasons, communication should never be command-based. The vehicle should communicate solely its intentions.

Communication can be directed or undirected. Pedestrians usually assume that any AVs communication is referring to themselves, hence using eHMIs with multiple pedestrians present has to be carried out in a way that minimizes mis-communication (i.e. either letting all pedestrians pass or not displaying a signal at all). Directed signalling minimizes this risk as other road users do not visually perceive the signal

of the eHMI. Dietrich *et al.* [57] found that pedestrians were not able to distinguish whether an undirected light signal was addressed to themselves or other traffic participants.Therefore, AVs should either use directed communication in ambiguous situations involving multiple pedestrians or no communication at all, as pedestrians will base their crossing decision on the approaching vehicles kinematics if no eHMI is present. The color of the visual eHMI stimulus may be of importance [216].

The most common eHMI display types are projection, high resolution displays and direct light. Semantics used include animations, concrete iconography, or text. For instance, Habibovic *et al.* developed a communication concept based on external light signals on the top of the windshield [86]. Using various light animations, the intention of the AV as well as the current driving mode such as 'I'm about to yield', 'I'm resting, and 'I'm about to start are displayed on the LED light bar. Clamann *et al.* [49] empirically examines similar models efficacy for giving signals to pedestrians. Further eHMI concepts include mimicking eye contacts by adding visible 'eyes to AVs – based on the well-known tendency for humans to perceive and design faces in cars– which can communicate detection and awareness of pedestrians through eye contact [42], as well as a virtual driver's mimicking furthermore facial expressions or hand signals. In addition to the pure visual-based communication between AVs and other TPs, some concepts also consider a combination of light and audio signals, as in the Google, Uber concepts and Mercedes-Benz concept car F015.

Most of the concepts presented here do not yet include detailed user studies and thus there remains a need for thorough evaluation including the behavioral and emotional responses of pedestrians in realistic environments. Different findings might be due to different eHMI concepts, diverse traffic scenarios, as well as different communication strategies. While research is still lacking in full understanding of the effects of eHMI on traffic, a large number of conceptual solutions have been proposed. Their influence on pedestrians, regarding their safety, experience and acceptance remains unclear. Most of these conceptual solutions are proposed by industry and involve some form of visual communication as the visual channel is the currently most used channel of communication in traffic as well as the best suited for communication at larger distances in busy environments.

C. Discussions

Game Theory has a long history of use in V2V (vehicle to vehicle) interactions in classical transport studies, as microscopic models underlying simulations of traffic flows and infrastructure design. Also multi-robot game theory systems are quite mature in robotics. These two streams have not generally been unified or applied to AV-pedestrian modelling, though this is beginning to emerge as an early research area. Like other sophisticated methods, game theoretic models can be computationally expensive.

Signalling models remain a distant research frontier. Physical actuators for eHMI signalling are currently being investigated by car manufacturers and recent years have seen

much patent activity in the area. But how to best use them to transmit information is not understood. There are currently no game-theoretic models using knowing and showing with explicit signalling but this would appear to be a fruitful area for future research. Eye contact is a particular form of signalling, but even in high level psychology research there remains an ongoing and lively debate about whether it is relevant or useful. The signalling methods reviewed here are mainly from qualitative studies, some work is still needed to implement their findings in algorithms for AVs.

V. EXPERIMENTAL RESOURCES

A. Pedestrian datasets

Large data sets are important resources for training and testing models at all levels, especially when they are annotated with 'ground truth information by humans. Their use has been common for low-level models such as detection and tracking, though there is currently a shortage of high quality annotated data for the higher-level models such as social interactions.

Major visual pedestrian datasets include the Caltech Pedestrian Benchmark [59], ETH [66], TUD-Brussels [218], Daimler [?], Stanford Drone Dataset [175], UCY Zara pedestrian dataset [124] and INRIA [51]. CityPersons [226] is a large dataset for pedestrian detection. Town Center Dataset [16] is a video dataset composed of 71.5k annotations.

Datasets used for pedestrian re-identification, i.e. having many images of the same people with identifiers include for example CUHK01 [126], CUHK02 [125] and CUHK03 [127], collected at a university campus and composed of thousands of bounding boxes of unique people. DUKEMTMC [173] and DUKEMTMC-reID [228] datasets have been developed in the Duke university campus and are used for tracking and re-identifying multiple people with multi-camera systems. MARKET-1501 [227] dataset provides 35k images of 1500 individuals but also comes with a 500k dataset of non-pedestrian street window distractors for training classifiers. Multi-Object Tracking Benchmark [143] collects diverse datasets and publishes new data. Several releases have already appeared: MOT15, MOT16 and MOT17.

PETA benchmark [52] is a mixture of several public datasets (e.g VIPER, SARC3D, PRID, MIT, I-LID, GRID, CAVIAR4REID, 3DPES), which has been used to recognize pedestrian attributes at far distance. The benchmark has been tested with an SVM method. Social ground truth annotations are much rarer. [36] and [37] collected high quality human annotations of physical and social events during pedestrian-vehicle interactions, including the presence and timings of the agents communicating with each other via eye contact, hand gestures, positions and speeds, and the final 'winners' of interactions which compete for road space during crossings.

Yang *et al.* [221] pointed out that in mixed urban scenarios, intelligent vehicles (IVs) have to cope with a certain number of surrounding pedestrians. Therefore, it is necessary to understand how vehicles and pedestrians interact with each other. They proposed a novel pedestrian trajectory dataset composed of CITR dataset and DUT dataset, so that the pedestrian motion models can be further calibrated and verified, especially

when the vehicle's influence on pedestrians plays an important role. In particular, the final trajectories of pedestrians and vehicles were refined by Kalman filters with linear point-mass model and nonlinear bicycle model, respectively, in which xy-velocity of pedestrians and longitudinal speed and orientation of vehicles were estimated.

Zhan et al. proposed INTERACTION dataset [225] which contains naturalistic motions of various traffic participants in a variety of highly interactive driving scenarios. Trajectory data was collected using drones and traffic cameras, containing data from multiple countries (USA, China, Germany and Bulgaria). There are four different driving scenarios, with their semantic maps provided: roundabouts, un-signalized intersection, signalized intersection, merging and lane changing. Chang *et al.* proposed Argoverse [43] containing two datasets and HD maps recorded from a self-driving car. Argoverse 3D Tracking is for 3D object annotations, it contains a collection of 11,052 tracks, and Argoverse Motion Forecasting is a curated collection of 324,557 scenarios, each 5 seconds long, for trajectory prediction. Each scenario contains the 2D, birds-eye-view centroid of each tracked object. ApolloScape dataset [214] was recorded in urban areas in China using various sensors. The dataset contains different road road users (vehicles, pedestrians, bicycles). The ApolloScape LeaderBoard shows the ranking and performance of the models tested on the dataset for different tasks, such as scene parsing, detection/tracking, trajectory prediction, self-localisation. The Intersection Drone (InD) dataset [25] contains naturalistic vehicle trajectories recorded using a drone at four German intersections. It provides the trajectories for thousands of road users and their types (e.g car, pedestrian, bicycle, truck), and can be used for example for road user prediction.

Person detection in off-road agricultural vehicle environments has become popular in recent years. Results from these studies are not well known in transport research but may transfer to on-carriageway and on-pavement AVs as they deal with similar types of pedestrian interactions. The National Robotics Engineering Center (NREC) Agricultural Person Detection Dataset [158] consists of labeled stereo video of people in orange and apple orchards taken from a tractor and a pickup truck, along with vehicle position data. The dataset combines a total of 76k labeled person images and 19k sampled person-free images. Gabriel *et al.* [73] present a dataset that focuses on action/intention recognition problems for human interactions with small robots in agriculture, including ten actors performing nine gestures and four activities. Stereo camera images, thermal camera images and Lidar point cloud data are recorded on grassland, under varying lighting conditions and distances. Kragh *et al.* [118] presents a multi-modal dataset for obstacle detection in agriculture containing 2h of raw sensor data from a tractor-mounted sensor system in a grass mowing scenario, including moving humans scattered in the field.

A summary of pedestrian datasets is given in the supplementary material Sect. **??** Table **??**.

B. Vehicle datasets

To train and test models of pedestrians interacting with vehicles, it is most likely useful to provide similar big data about vehicles as well as about pedestrians. This may include ground truth information on vehicle location and motion, but also high level social annotations to use in studies of interaction with pedestrians. Visual data available includes the Berkeley DeepDrive Video (BDDV) dataset [220], currently the largest vehicle dataset publicly available with 10k hours of driving videos around the world. KITTI dataset [75] provides a one hour video of a vehicle driving in an urban environment. Caesar *et al.* [32] presented *nuScenes* a dataset for autonomous driving composed of multiple sensor data (RGB, LIDAR, RADAR) from two cities and containing 1k scenes. A summary of vehicle datasets is given in the supplementary material Sec. **??** Table **??**.

C. Pedestrian and Driving Simulators

Three types of vehicle and pedestrian simulation research work exist, these categories are listed below, some concrete examples will follow in the next paragraphs. The design and source code for simulators like the ones introduced below are often not made publicly available, thus making it difficult for others researchers to investigate and replicate experiments, hence there is a clear need for more open-source simulators. A summary of the simulators is proposed in the supplementary material Sec. **??** Table **??**.

a) Vehicle-Pedestrian Simulators: Micro or macro simulations model both vehicle and pedestrian behavior. Most of these simulations rely on sets of behavioral rules for both agents. These simulators are primarily used for road design purposes and for policy decisions such as the cellular automata-based simulators proposed in [71] and [130] where vehicle-pedestrian crossing behaviour is studied at crosswalks. Feliciani *et al.* [71] further evaluated the necessity of introducing a new crosswalk and/or switching to a traffic light. Chao *et al.* [44] developed a microscopic-based traffic simulator based on a force model to represent the behaviour and interactions between the road users, and aimed for autonomous vehicle development and testing. Chen *et al.* [45] proposed a simulation platform composed of several behaviour models at crosswalks for vehicle-pedestrian conflicts assessment. Gupta *et al.* [83] developed a simulation model, using Matlab and the open-source SUMO (Simulation of Urban Mobility), for autonomous vehicle-pedestrian negotiations at unmarked intersections, considering different pedestrian behaviours. VirtuoCity is an example of physical vehicle-pedestrian simulators. It is composed of a pedestrian simulator, HIKER [181], which is a virtual reality 'CAVE-based environment for pedestrian behavior analysis, a driving simulator [106] and a truck simulator for driver behaviour understanding. IFSTTAR [98] also possesses a pedestrian simulator and developed a driving simulator for driver behavior analysis and human-machine interactions, an immersive simulator for cars, motorcycles and pedestrians behavior simulation, a driving simulator with human assistive devices and a bicycle simulator. Commercial

products include STEPS [146] software for and Legion [123] simulating pedestrian dynamics.

b) Vehicle Simulators: Vehicle simulators are physical platforms where drivers encounter virtual pedestrians (dummies) in order to study driver yielding behaviors in specific interaction scenarios. Simulators such as [149] studied driver-pedestrian interactions in mixed traffic environments using a driving simulator (DriveSafetys DS-600c Research Simulator). JARI-ARV (Augmented Reality Vehicles) [107] is a road running driving simulator and JARI-OVDS (Omnidirectional View Driving Simulator is a driving simulator with 360-degree spherical screen and a rocking device. The University of Iowa [205] has developed a driving simulator. A previous review on driving simulators is presented in [192]..

c) Pedestrian Simulators: Pedestrian simulators are VR (Virtual Reality) based environments where pedestrian participants encounter virtual vehicles in order to study pedestrian perception and decision making subject to various oncoming vehicle behaviors [184]. For example, Camara *et al.* [34], [35] developed their experiments with the free version of Unity game engine[2], 3D models were imported from the Asset Store, and used a HTC Vive VR headset for pedestrians interacting with a game theoretic autonomous vehicle. Their results showed that VR is a reliable setup for measuring human behaviour for the development and testing of AV technology. Mahadevan *et al.* [135] presented OnFoot, a VR pedestrian simulator that studies pedestrian interactions with autonomous vehicles in a mixed traffic environment. The Technical University of Munich also developed a pedestrian simulator [70] composed of a head-mounted display, a motion capture system and a driving simulator software. This setup could be connected to a driving simulator enabling multi-agent studies while extracting the participants gait during the crossing process. The current setup utilizes the Unity 3D game engine (with a VIVE HMD) and is sometimes coupled with VIVE Trackers for a virtual self-representation to create an immersive virtual environment enabling fast implementations and evaluation of eHMI concepts [57]. PedSim [77] is a free crowd simulation software.

VI. Conclusions

Pedestrian sensing, detection and kinematic tracking are now well understood and have mature models as reviewed in Part I [33]. Moving from simple kinematic tracking and prediction of pedestrian motions can however depend on extremely high-level models of the state transition required by tracking and prediction. Going far beyond simple random velocity walk models, the present review has shown that there is much scope here to integrate models of pedestrians as intelligent, goal-based, psychological, active, and interactive agents at several levels.

Unlike the more mature methods reviewed in Part I, this review does not recommend particular software implementations for algorithms at these levels, because they remain active research areas rather than completed and standardizable tools. This review finds that many conceptual issues first need to be

cleared, before mathematical interfaces – such as probabilities – can be created to link models at these layers, and only then standardized software development can become a reality. (The only exception to this would be for entirely end-to-end machine learning systems, which are not generally considered to be safe or practical due to their lack of transparency.)

At the level of single pedestrian modelling, there now exist good control theoretic models of optimal walking behaviour from known origin to known destination. Here, pedestrians do not usually walk in straight lines, but optimise gradual turning during walking to move in smooth curves. There has been some recent research success in inferring likely destinations from historical data and partial trajectories.

When interaction with other agents is included, models of pedestrians rapidly become more complex and much less well understood. Suboptimal models include only finite orders of epistemological models of pedestrians beliefs, raising the open question of how to handle higher order beliefs about beliefs. Recent game theory approaches have just begun to find optimal behaviours in these higher-order belief cases but only under various simplifying assumptions.

There has been a general shift away from psychology-informed models, using empirical findings such as demographics predicting behaviours, to purely big-data-driven models which learn aspects of such theories internally as black boxes, usually aiming only to predict the behaviour rather than give theoretical explanations of it.

The role of signalling between pedestrians and vehicles during interactions has been studied qualitatively, but is not yet understood at the algorithmic level. Psychologists and road safety designers have evaluated and commercialised many signalling mechanisms, such as flashing of headlights, use of horns, and custom communication light signals. Finding algorithmic strategies to make optimal use of them, and to process information from receiving signals from others, suitable for real-time AV control, remains an open and important question.

References

[1] I. Ajzen. Attitudes and personality traits. *Attitudes, Personality, and Behavior*, pages 2–24, 1988.

[2] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proc. of IEEE CVPR*, pages 961–971, 2016.

[3] W. K. Alhajyaseen and M. Iryo-Asano. Studying critical pedestrian behavioral changes for the safety assessment at signalized crosswalks. *Safety Science*, 91:351 – 360, 2017.

[4] S. Ali and M. Shah. A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proc. of IEEE CVPR*, pages 1–6, 2007.

[5] M. Althoff. *Reachability analysis and its application to the safety assessment of autonomous cars*. PhD thesis, Technische Universität München, 2010.

[6] J. Amirian, J.-B. Hayet, and J. Pettré. Social ways: Learning multi-modal distributions of pedestrian trajectories with gans. In *Proc. of IEEE CVPR Workshops*, 2019.

[7] J. Andersson, A. Habibovic, M. Klingegård, C. Englund, and V. Malmsten-Lundgren. Hello Human, can you read my mind? *ERCIM News*, (109):36–37, 2017.

[8] G. Antonini, M. Bierlaire, and M. Weber. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8):667 – 687, 2006.

[9] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz. On the nonholonomic nature of human locomotion. *Autonomous Robots*, 25(1-2):25–35, 2008.

---

[2]https://unity.com/

[10] E. Avineri, D. Shinar, and Y. O. Susilo. Pedestrians behaviour in cross walks: The effects of fear of falling and age. *Accident Analysis & Prevention*, 44(1):30 – 34, 2012.

[11] H. Bai, S. Cai, N. Ye, D. F. C. Hsu, and W. S. Lee. Intention-aware online pomdp planning for autonomous driving in a crowd. *Proc. of IEEE ICRA*, pages 454–460, 2015.

[12] M. Batty. Predicting where we walk. *Nature*, 388(6637):1920, 1997.

[13] D. Bauer. Comparing pedestrian movement simulation models for a crossing area based on real world data. In *Pedestrian and Evacuation Dynamics*, pages 547–556. Springer, 2011.

[14] D. Bauer, S. Seer, and N. Brändle. Macroscopic pedestrian flow simulation for designing crowd control measures in public transport after special events. In *Proc. of the Summer Computer Simulation Conference*, pages 1035–1042, 2007.

[15] N. Bellomo and C. Dogbe. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM review*, 53(3):409–463, 2011.

[16] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *Proc. of IEEE CVPR*, pages 3457–3464, 2011.

[17] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun. Learning motion patterns of people for compliant robot motion. *International Journal of Robotics Research*, 24(1):31–48, 2005.

[18] M. Bennewitz, W. Burgard, and S. Thrun. Learning motion patterns of persons for mobile service robots. In *Proc. of IEEE ICRA*, volume 4, pages 3601–3606, 2002.

[19] A. Bera, S. Kim, T. Randhavane, S. Pratapa, and D. Manocha. GLMP - realtime pedestrian path prediction using global and local movement patterns. In *Proc. of IEEE ICRA*, pages 5528–5535, 2016.

[20] A. Bera and D. Manocha. Pedlearn: Realtime pedestrian tracking, behavior learning, and navigation for autonomous vehicles. In *Proc. of IROS 9th International workshop on Planning, Perception and Navigation for Intelligent Vehicles*, 2017.

[21] I. M. Bernhoft and G. Carstensen. Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(2):83 – 95, 2008.

[22] A. Bhattacharyya, M. Fritz, and B. Schiele. Long-term on-board prediction of pedestrians in traffic scenes. In *Proc. of 1st Conference on Robot Learning*, 2017.

[23] V. J. Blue and J. L. Adler. Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B: Methodological*, 35(3):293 – 312, 2001.

[24] J. Bock, T. Beemelmanns, M. Klösges, and J. Kotte. Self-learning trajectory prediction with recurrent neural networks at intelligent intersections. In *Proc. of VEHITS*, 2017.

[25] J. Bock, R. Krajewski, T. Moers, L. Vater, S. Runde, and L. Eckstein. The ind dataset: A drone dataset of naturalistic vehicle trajectories at german intersections. *arXiv preprint arXiv:1911.07602*, 2019.

[26] S. Bonneaud and W. H. Warren. A behavioral dynamics approach to modeling realistic pedestrian behavior. In *Proc. of 6th International Conference on Pedestrian and Evacuation Dynamics*, pages 1–14, 2012.

[27] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmuedderich. Pedestrian crossing prediction using multiple context-based models. In *Proc. of IEEE ITSC*, pages 378–385, 2014.

[28] A. Borgers, A. Kemperman, and H. Timmermans. *Modeling Pedestrian Movement in Shopping Street Segments*, chapter Chapter 5, pages 87–111. 2009.

[29] A. Borgers and H. Timmermans. A model of pedestrian route choice and demand for retail facilities within innercity shopping areas. *Geographical Analysis*, 18(2):115–128, 2010.

[30] R. Brooks. The big problem with self-driving cars is people and we'll go out of our way to make the problem worse. *IEEE Spectrum*, 2017.

[31] F. Broz, I. Nourbakhsh, and R. Simmons. Planning for human-robot interaction using time-state aggregated pomdps. In *Proc. of the 23rd National Conference on Artificial Intelligence*, volume 3, pages 1339–1344, 2008.

[32] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuScenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027*, 2019.

[33] F. Camara, N. Bellotto, S. Cosar, D. Nathanael, M. Althoff, J. Wu, J. Ruenz, A. Dietrich, and C. W. Fox. Pedestrian models for autonomous driving Part I: low level models, from sensing to tracking. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[34] F. Camara, P. Dickinson, N. Merat, and C. W. Fox. Towards game theoretic AV controllers:measuring pedestrian behaviour in virtual reality. In *Proc. of IEEE/RSJ IROS Workshops*, 2019.

[35] F. Camara, P. Dickinson, N. Merat, and C. W. Fox. Examining pedestrian behaviour in virtual reality. In *Transport Research Arena (TRA) (Conference canceled)*, 2020.

[36] F. Camara, O. Giles, R. Madigan, M. Rothmüller, P. Holm Rasmussen, S. A. Vendelbo-Larsen, G. Markkula, Y. M. Lee, L. Garach, N. Merat, and C. W. Fox. Filtration analysis of pedestrian-vehicle interactions for autonomous vehicles control. In *Proc. of the International Conference on Intelligent Autonomous Systems (IAS-15) Workshops*, 2018.

[37] F. Camara, O. Giles, R. Madigan, M. Rothmller, P. H. Rasmussen, S. A. Vendelbo-Larsen, G. Markkula, Y. M. Lee, L. Garach, N. Merat, and C. W. Fox. Predicting pedestrian road-crossing assertiveness for autonomous vehicle control. In *Proc. of IEEE ITSC*, pages 2098–2103, 2018.

[38] F. Camara, N. Merat, and C. W. Fox. A heuristic model for pedestrian intention estimation. In *Proc. of IEEE ITSC*, pages 3708–3713, 2019.

[39] F. Camara, R. Romano, G. Markkula, R. Madigan, N. Merat, and C. Fox. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Proc. of Measuring Behavior*, 2018.

[40] C. Caramuta, G. Collodel, C. Giacomini, C. Gruden, G. Longo, and P. Piccolotto. Survey of detection techniques, mathematical models and simulation software in pedestrian dynamics. *Transportation Research Procedia*, 25(Supplement C):551 – 567, 2017.

[41] D. Castelvecchi. Can we open the black box of AI? *Nature News*, 538(7623):20, 2016.

[42] C.-M. Chang, K. Toda, D. Sakamoto, and T. Igarashi. Eyes on a Car: An Interface Design for Communication Between an Autonomous Car and a Pedestrian. In *Proc. of ACM AutomotiveUI*, pages 65–73, 2017.

[43] M.-F. Chang, J. W. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays. Argoverse: 3d tracking and forecasting with rich maps. In *Proc. of IEEE CVPR*, 2019.

[44] Q. Chao, X. Jin, H.-W. Huang, S. Foong, L.-F. Yu, and S.-K. Yeung. Force-based heterogeneous traffic simulation for autonomous vehicle testing. In *Proc. of ICRA*, pages 8298–8304. IEEE, 2019.

[45] P. Chen, W. Zeng, and G. Yu. Assessing right-turning vehicle-pedestrian conflicts at intersections using an integrated microscopic simulation model. *Accident Analysis & Prevention*, 129:211–224, 2019.

[46] Z. Chen, C. Wu, N. Lyu, G. Liu, and Y. He. Pedestrian-vehicular collision avoidance based on vision system. In *Proc. of IEEE ITSC*, pages 11–15, 2014.

[47] B. Cheng, X. Xu, Y. Zeng, J. Ren, and S. Jung. Pedestrian trajectory prediction via the social-grid LSTM model. *The Journal of Engineering*, 2018(16):1468–1474, 2018.

[48] H. Choset, K. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. Kavraki, and S. Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press, 2005.

[49] M. Clamann, M. Aubert, and M. L. Cummings. Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles. In *Proc. of TRB*, 2017.

[50] A. Cosgun, D. A. Florencio, and H. I. Christensen. Autonomous person following for telepresence robots. In *Proc. of IEEE ICRA*, pages 4335–4342, 2013.

[51] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE CVPR*, volume 1, pages 886–893 vol. 1, 2005.

[52] Y. Deng, P. Luo, C. C. Loy, and X. Tang. Pedestrian attribute recognition at far distance. In *Proc. of the 22nd ACM international Conference on Multimedia*, pages 789–792. ACM, 2014.

[53] N. Deo and M. M. Trivedi. Learning and predicting on-road pedestrian behavior around vehicles. In *Proc. of IEEE ITSC*, pages 1–6, 2017.

[54] N. Deo and M. M. Trivedi. Trajectory forecasts in unknown environments conditioned on grid-based plans. *arXiv preprint arXiv:2001.00735*, 2020.

[55] C. Dias, M. Abdullah, M. Sarvi, R. Lovreglio, and W. Alhajyaseen. Modeling and simulation of pedestrian movement planning around corners. *Sustainability*, 11(19):5501, 2019.

[56] P. Dickinson, K. Gerling, K. Hicks, J. Murray, J. Shearer, and S. Greenwood. Virtual reality crowd simulation: effects of agent density on user experience and behaviour. *Virtual Reality*, 2018.

[57] A. Dietrich, J.-H. Willrodt, K. Wagner, and K. Bengler. Projection-based external human machine interfaces enabling interaction between automated vehicles and pedestrians. In *Driving Simulation Conference*, 2018.

[58] J. Dijkstra, H. J. Timmermans, and A. Jessurun. A multi-agent cellular automata system for visualising simulated pedestrian activity. In *Theory and Practical Issues on Cellular Automata*, pages 29–36. Springer, 2001.

[59] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *Proc. of CVPR*, 2009.

[60] C. Dondrup, N. Bellotto, M. Hanheide, K. Eder, and U. Leonards. A computational model of human-robot spatial interactions based on a qualitative trajectory calculus. *Robotics*, 4(1):63–102, 2015.

[61] C. Dondrup, N. Bellotto, F. Jovan, and M. Hanheide. Real-time multisensor people tracking for human-robot spatial interaction. In *Proc. of ICRA Workshop on Machine Learning for Social Robotics*, 2015.

[62] P. Duckworth, M. Al-Omari, J. Charles, D. C. Hogg, and A. G. Cohn. Latent dirichlet allocation for unsupervised activity analysis on an autonomous mobile robot. In *Proc. of AAAI*, 2017.

[63] P. Duckworth, Y. Gatsoulis, F. Jovan, N. Hawes, D. C. Hogg, and A. G. Cohn. Unsupervised learning of qualitative motion behaviours by a mobile robot. In *Proc. of AAMAS*, pages 1043–1051, 2016.

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

[65] R. Elvik. A review of game-theoretic models of road user behaviour. *Accident Analysis & Prevention*, 62:388 – 396, 2014.

[66] A. Ess, B. Leibe, and L. V. Gool. Depth and appearance for mobile scene analysis. In *Proc. of IEEE ICCV*, pages 1–8, 2007.

[67] R. Etikyala, S. Göttlich, A. Klar, and S. Tiwari. Particle methods for pedestrian flow models: from microscopic to nonlocal continuum models. *Mathematical Models and Methods in Applied Sciences*, 24(12):2503–2523, 2014.

[68] D. Evans and P. Norman. Predicting adolescent pedestrians road-crossing intentions: an application and extension of the theory of planned behaviour. *Health Education Research*, 18(3):267–277, 2003.

[69] B. R. Fajen and W. H. Warren. Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):343, 2003.

[70] I. Feldstein, A. Dietrich, S. Milinkovic, and K. Bengler. A pedestrian simulator for urban crossing scenarios. volume 49, pages 239 – 244, 2016. Proc. of IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems.

[71] C. Feliciani, L. Crociani, A. Gorrini, G. Vizzari, S. Bandini, and K. Nishinari. A simulation model for non-signalized pedestrian crosswalks based on evidence from on field observation. *Intelligenza Artificiale*, 11(2):117–138, 2017.

[72] C. W. Fox, F. Camara, G. Markkula, R. Romano, R. Madigan, and N. Merat. When should the chicken cross the road?: Game theory for autonomous vehicle - human interactions. In *Proc. of VEHITS*, 2018.

[73] A. Gabriel, S. Cosar, N. Bellotto, and P. Baxter. A dataset for action recognition in the wild. In *Proc. of TAROS*, pages 362–374, 2019.

[74] M. Garzón, D. Garzón-Ramos, A. Barrientos, and J. d. Cerro. Pedestrian trajectory prediction in large infrastructures. In *Proc. of International Conference on Informatics in Control, Automation and Robotics*, pages 381–389, 2016.

[75] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[76] P. Gipps and B. Marksjö. A micro-simulation model for pedestrian flows. *Mathematics and Computers in Simulation*, 27(2):95 – 105, 1985.

[77] C. Gloor. Pedsim : Pedestrian crowd simulator, 2012.

[78] M. Goldhammer, S. Khler, K. Doll, and B. Sick. Camera based pedestrian path prediction by means of polynomial least-squares approximation and multilayer perceptron neural networks. In *Proc. of SAI Intelligent Systems Conference (IntelliSys)*, pages 390–399, 2015.

[79] A. Gorrini, S. Bandini, and M. Sarvi. Group dynamics in pedestrian crowds: Estimating proxemic behavior. *Transportation Research Record*, 2421(1):51–56, 2014.

[80] A. Gorrini, L. Crociani, G. Vizzari, and S. Bandini. Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 59:269–285, 2018.

[81] D. Gunning. Explainable artificial intelligence (XAI). *Defense Advanced Research Projects Agency (DARPA)*, 2, 2017.

[82] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proc. of IEEE CVPR*, pages 2255–2264, 2018.

[83] S. Gupta, M. Vasardani, and S. Winter. Negotiation between vehicles and pedestrians for the right of way at intersections. *IEEE Transactions on Intelligent Transportation Systems*, 20(3):888–899, 2018.

[84] N. Guguen, S. Meineri, and C. Eyssartier. A pedestrians stare and drivers stopping behavior: A field experiment at the pedestrian crossing. *Safety Science*, 75:87 – 89, 2015.

[85] G. Habibi, N. Jaipuria, and J. P. How. Context-aware pedestrian motion prediction in urban intersections. In *Proc. of IEEE ITSC*, 2018.

[86] A. Habibovic, V. M. Lundgren, J. Andersson, M. Klingegård, T. Lagström, A. Sirkka, J. Fagerlönn, C. Edgren, R. Fredriksson, S. Krupenia, D. Saluäär, and P. Larsson. Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology*, 9:1336, 2018.

[87] E. T. Hall. A system for the notation of proxemic behavior. *American Anthropologist*, 65(5):1003–1026, 1963.

[88] E. T. Hall. The hidden dimension. 1969.

[89] W. A. Harrell. The impact of pedestrian visibility and assertiveness on motorist yielding. *The Journal of Social Psychology*, 133(3):353–360, 1993.

[90] Helbing and Molnár. Social force model for pedestrian dynamics. *Physical Review E*, 51:4282–4286, 1995.

[91] P. Henry, C. Vollmer, B. Ferris, and D. Fox. Learning to navigate through crowded environments. In *Proc. of IEEE ICRA*, pages 981–986, 2010.

[92] C. Holland and R. Hill. The effect of age, gender and driver status on pedestrians intentions to cross the road in risky situations. *Accident Analysis & Prevention*, 39(2):224 – 237, 2007.

[93] C. Holland and R. Hill. Gender differences in factors predicting unsafe crossing decisions in adult pedestrians across the lifespan: A simulation study. *Accident Analysis and Prevention*, 42(4):1097 – 1106, 2010.

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

[95] S. Hoogendoorn and P. HL Bovy. Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications and Methods*, 24(3):153–172, 2003.

[96] R. Hug, S. Becker, W. Hübner, and M. Arens. Particle-based pedestrian path prediction using LSTM-MDL models. *CoRR*, abs/1804.05546, 2018.

[97] R. L. Hughes. A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological*, 36(6):507 – 535, 2002.

[98] IFSTARR. Simulators, 2018. http://www.ifsttar.fr/en/exceptional-facilities/simulators/.

[99] T. Ikeda, Y. Chigodo, D. Rea, F. Zanlungo, M. Shiomi, and T. Kanda. Modeling and prediction of pedestrian behavior based on the sub-goal concept. In *Proc. of Robotics: Science and Systems*, 2012.

[100] I. Institute for Advanced Motorists. *How to Be A Better Driver: Advanced Driving the Essential Guide*. Motorbooks, 2007.

[101] S. International. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, 2018.

[102] M. Iryo-Asano and W. K. Alhajyaseen. Analysis of pedestrian clearance time at signalized crosswalks in Japan. *Procedia Computer Science*, 32:301 – 308, 2014.

[103] M. Iryo-Asano and W. K. Alhajyaseen. Modeling pedestrian crossing speed profiles considering speed change behavior for the safety assessment of signalized intersections. *Accident Analysis & Prevention*, 108:332 – 342, 2017.

[104] M. Iryo-Asano, W. K. M. Alhajyaseen, and H. Nakamura. Analysis and modeling of pedestrian crossing behavior during the pedestrian flashing green interval. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):958–969, 2015.

[105] R. Isaacs. Games of pursuit. 1951.

[106] A. Jamson, A. Horrobin, and R. Auckland. Whatever happened to the lads? design and development of the new University of Leeds driving simulator. In *Proc. of Driving Simulator Conference*, 2007.

[107] JARI. Simulators, 2018. http://www.jari.or.jp/.

[108] F. T. Johora and J. Müller. Modeling interactions of multimodal road users in shared spaces. In *Proc. of IEEE ITSC*, 2018.

[109] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto. Intent-aware long-term prediction of pedestrian motion. In *Proc. of IEEE ICRA*, pages 2543–2549, 2016.

[110] A. Katz, D. Zaidel, and A. Elgrishi. An experimental study of driver and pedestrian interaction during the crossing conflict. *Human Factors*, 17(5):514–527, 1975.

[111] K. Kawamoto, Y. Tomura, and K. Okamoto. Learning pedestrian dynamics with kriging. In *Proceedings IEEE/ACIS International Conference on Computer and Information Science*, pages 1–4, 2016.

[112] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert. Activity forecasting. In *Proc. of ECCV*, pages 201–214, 2012.

[113] K. Kitazawa and M. Batty. Pedestrian behaviour modelling. In *Proc. of the International Conference on Developments in Design and Decision Support Systems in Architecture and Urban Planning*, 2004.

[114] J. Kober, J. A. Bagnell, and J. Peters. Reinforcement learning in robotics: A survey. *International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[115] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila. Context-based pedestrian path prediction. In *Proc. of ECCV*, pages 618–633, 2014.

[116] M. Koschi, C. Pek, M. Beikirch, and M. Althoff. Set-based prediction of pedestrians in urban environments considering formalized traffic rules. In *Proc. of IEEE ITSC*, 2018.

[117] P. Kothari and A. Alahi. Adversarial loss for human trajectory prediction. In *Proc. of hEART*, 2019.

[118] M. Kragh, P. Christiansen, M. Laursen, M. Larsen, K. Steen, O. Green, H. Karstoft, and R. Jrgensen. Fieldsafe: Dataset for obstacle detection in agriculture. *Sensors*, 17, 2017.

[119] H. Kretzschmar, M. Kuderer, and W. Burgard. Learning to predict trajectories of cooperatively navigating agents. In *Proc. of IEEE ICRA*, pages 4015–4020, 2014.

[120] E. Kruse, R. Gutsche, and F. M. Wahl. Acquisition of statistical motion patterns in dynamic environments and their application to mobile robot motion planning. In *Proc. of IEEE/RSJ IROS*, volume 2, pages 712–717 vol.2, 1997.

[121] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. In *Proc. of IEEE CVPR*, pages 336–345, 2017.

[122] S. Lefèvre, D. Vasquez, and C. Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, 1(1):1, 2014.

[123] Legion. Legion for aisun, 1997. http://www.legion.com/legion-for-aimsun-%E2%80%93-vehicle-pedestrian-simulation.

[124] A. Lerner, Y. Chrysanthou, and D. Lischinski. Crowds by example. In *Computer Graphics Forum*, volume 26, pages 655–664. Wiley Online Library, 2007.

[125] W. Li and X. Wang. Locally aligned feature transforms across views. In *Proc. of IEEE CVPR*, pages 3594–3601, 2013.

[126] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *Proc. of Asian Conference on Computer Vision*, pages 31–44. Springer, 2012.

[127] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proc. of IEEE CVPR*, pages 152–159, 2014.

[128] C. Lichtenthäler and A. Kirsch. Towards legible robot navigation - how to increase the intend expressiveness of robot navigation behavior. In *Int. Conf. Soc. Robot. Embodied Commun. Goals Intention*, 2013.

[129] R. Lobjois and V. Cavallo. The effects of aging on street-crossing behavior: from estimation to actual crossing. *Accident Analysis & Prevention*, 41(2):259–267, 2009.

[130] L. Lu, G. Ren, W. Wang, C.-Y. Chan, and J. Wang. A cellular automaton simulation model for pedestrian and vehicle interaction behaviors at unsignalized mid-block crosswalks. *Accident Analysis & Prevention*, 95:425–437, 2016.

[131] V. M. Lundgren, A. Habibovic, J. Andersson, T. Lagström, M. Nilsson, A. Sirkka, J. Fagerlönn, R. Fredriksson, C. Edgren, S. Krupenia, and D. Saluäär. Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context? In *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2016 International Conference on Human Factors in Transportation*, pages 485–497. Springer International Publishing, 2017.

[132] W. Luo, X. Zhao, and T. Kim. Multiple object tracking: A review. *CoRR*, abs/1409.7618, 2014.

[133] G. G. Lvs. Modeling and simulation of pedestrian traffic flow. *Transportation Research Part B: Methodological*, 28(6):429 – 443, 1994.

[134] W. Ma, D. Huang, N. Lee, and K. M. Kitani. Forecasting interactive dynamics of pedestrians with fictitious play. In *Proc. of IEEE CVPR*, 2017.

[135] K. Mahadevan, E. Sanoubari, S. Somanath, J. E. Young, and E. Sharlin. AV-Pedestrian interaction design using a pedestrian mixed traffic simulator. In *Proc. of the Conference on Designing Interactive Systems*, pages 475–486. ACM, 2019.

[136] K. Mahadevan, S. Somanath, and E. Sharlin. Can Interfaces Facilitate Communication in Autonomous Vehicle-Pedestrian Interaction? In *Proc. of ACM/IEEE HRI*, 2018.

[137] K. Mahadevan, S. Somanath, and E. Sharlin. Communicating awareness and intent in autonomous vehicle-pedestrian interaction. In *Proc. of the CHI Conference on Human Factors in Computing Systems*, page 429. ACM, 2018.

[138] L. Manenti, S. Manzoni, G. Vizzari, K. Ohtsuka, and K. Shimura. Towards an agent-based proxemic model for pedestrian and group dynamic. In *Proc. of the 11th WOA Workshop*, volume 621, 2010.

[139] G. Markkula, R. Romano, R. Madigan, C. W. Fox, O. T. Giles, and N. Merat. Models of human decision-making as tools for estimating and optimizing impacts of vehicle automation. *Transportation Research Record*, 2018.

[140] A. Martin. Interactive motion prediction using game theory. Master's thesis, 2013. Technical University of Munich.

[141] M. Matthews, G. Chowdhary, and E. Kieson. Intent communication between autonomous vehicles and pedestrians. *CoRR*, abs/1708.07123, 2017.

[142] B. L. Mesmer and C. L. Bloebaum. Modeling decision and game theory based pedestrian velocity vector decisions with interacting individuals. *Safety Science*, 87:116 – 130, 2016.

[143] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler. MOT16: A benchmark for multi-object tracking. *CoRR*, abs/1603.00831, 2016.

[144] A. Mögelmose, M. M. Trivedi, and T. B. Moeslund. Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations. In *Proc. of IEEE IV*, pages 330–335, 2015.

[145] B. T. Morris and M. M. Trivedi. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1114–1127, 2008.

[146] Mott-Macdonald. Steps, simulation pedestrian dynamics, 1997. https://www.stepsmottmac.com/.

[147] M. Moussaïd, D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proc. of the National Academy of Sciences*, 108(17):6884–6888, 2011.

[148] D. Nathanael, E. Portouli, V. Papakostopoulos, K. Gkikas, and A. Amditis. Naturalistic observation of interactions between car drivers and pedestrians in high density urban settings. In *Proc. of the 20th Congress of the International Ergonomics Association*, pages 389–397, 2019.

[149] H. Obeid, H. Abkarian, M. Abou-Zeid, and I. Kaysi. Analyzing driver-pedestrian interaction in a mixed-street environment using a driving simulator. *Accident Analysis & Prevention*, 108:56–65, 2017.

[150] S. Okazaki. A study of pedestrian movement in architectural space, part 1: Pedestrian movement by the application on of magnetic models. *Trans. AIJ*, 283:111–119, 1979.

[151] A. Olivier, J. Bruneau, R. Kulpa, and J. Pettr. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments. *IEEE Transactions on Visualization and Computer Graphics*, 24(7):2251–2263, 2018.

[152] J. Oxley, B. Fildes, E. Ihsen, R. Day, and J. Charlton. An investigation of road crossing behaviour of older pedestrians. *Transportation Research Board*, 1995.

[153] E. Papadimitriou, G. Yannis, and J. Golias. A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3):242 – 255, 2009.

[154] A. V. Papadopoulos, L. Bascetta, and G. Ferretti. Generation of human walking paths. In *Proc. of IEEE/RSJ IROS*, pages 1676–1681, 2013.

[155] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. *Proc. of IEEE ICCV*, pages 261–268, 2009.

[156] S. Pellegrini, A. Ess, and L. Van Gool. *Predicting Pedestrian Trajectories*, pages 473–491. Springer London, 2011.

[157] T. Petzoldt, K. Schleinitz, and R. Banse. Laboruntersuchung zur potenziellen Sicherheitswirkung einer vorderen Bremsleuchte in PKW. *Zeitschrift für Verkehrssicherheit*, 63(1):19–24, 2017.

[158] Z. Pezzementi, T. Tabor, P. Hu, J. K. Chang, D. Ramanan, C. Wellington, B. P. Wisely Babu, and H. Herman. Comparing apples and oranges: Off-road pedestrian detection on the national robotics engineering center agricultural person-detection dataset. *Journal of Field Robotics*, 35(4):545–563, 2018.

[159] E. Portouli, D. Nathanael, K. Gkikas, A. Amditis, and P. Loizos. Field observations of interactions among drivers at unsignalized urban intersections. In *Proc. of the 20th Congress of the International Ergonomics Association*, 2019.

[160] A. Puydupin-Jamin, M. Johnson, and T. Bretl. A convex approach to inverse optimal control and its application to modeling human locomotion. In *Poc. of IEEE ICRA*, pages 531–536, 2012.

[161] R. Quintero, I. P. Alonso, D. Fernández-Llorca, and M. Sotelo. Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12, 2018.

[162] R. Quintero, I. Parra, D. F. Llorca, and M. Sotelo. Pedestrian intention and pose prediction through dynamical models and behaviour classification. In *Proc. of IEEE ITSC*, pages 83–88, 2015.

[163] N. Radwan, A. Valada, and W. Burgard. Multimodal interaction-aware motion prediction for autonomous street crossing. *arXiv preprint arXiv:1808.06887*, 2018.

[164] H. A. Rakha, I. Zohdy, and R. K. Kamalanathsharma. Agent-based game theory modeling for driverless vehicles at intersections. 2013.

[165] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[166] A. Rasouli, I. Kotseruba, and J. K. Tsotsos. Agreeing to cross: How drivers and pedestrians communicate. *Proc. of IEEE IV*, pages 264–269, 2017.

[167] A. Rasouli and J. K. Tsotsos. Joint attention in driver-pedestrian interaction: from theory to practice. *CoRR*, abs/1802.02522, 2018.

[168] A. Rasouli and J. K. Tsotsos. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, 21(3):900–918, 2020.

[169] E. Rehder and H. Kloeden. Goal-directed pedestrian prediction. *Proc. of IEEE ICCV Workshop*, pages 139–147, 2015.

[170] E. Rehder, F. Wirth, M. Lauer, and C. Stiller. Pedestrian prediction by planning using deep neural networks. In *Proc. of IEEE ICRA*, pages 5903–5908, 2018.

[171] D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf. A literature review on the prediction of pedestrian behavior in urban scenarios. In *Proc. of IEEE ITSC*, pages 3105–3112, 2018.

[172] J. Rios-Martinez, A. Spalanzani, and C. Laugier. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics*, 7(2):137–153, 2015.

[173] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi. Performance measures and a data set formulti-target, multi-camera tracking. In *Proc. of ECCV Workshops*, pages 17–35, 2016.

[174] M. Risto, C. Emmenegger, E. Vinkhuyzen, M. Cefkin, and J. Hollan. Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. In *Proc. of the 9th International Driving Symposium on Human Factors in Driver Assessment*, pages 186–192, 2017.

[175] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Proc. of ECCV*, pages 549–565, 2016.

[176] N. Ronald, L. Sterling, and M. Kirley. A conceptual framework for specifying and developing pedestrian models. 2020.

[177] D. Rothenbücher, J. Li, D. Sirkin, B. Mok, and W. Ju. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *Proc. of IEEE RO-MAN*, pages 795–802, 2016.

[178] A. Rudenko, L. Palmieri, and K. O. Arras. Joint long-term prediction of human motion using a planning-based social force approach. In *Proc. of IEEE ICRA*, 2018.

[179] C. Rudin-Brown and S. Jamson. *Behavioural adaptation and road safety: Theory, evidence and action*. CRC Press, 2013.

[180] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In *Proc. of IEEE CVPR*, pages 1349–1358, 2019.

[181] E. Sadraei, R. Romano, N. Merat, J. Pedro, Y. M. Lee, R. Madigan, C. Uzondu, W. Lyu, and A. Tomlinson. Vehicle-pedestrian interaction: A distributed simulation study. In *Proc. of Driving Simulation Conference*, 2020.

[182] A. Schadschneider, W. Klingsch, H. Klüpfel, T. Kretz, C. Rogsch, and A. Seyfried. *Evacuation Dynamics: Empirical Results, Modeling and Applications*, pages 3142–3176. Springer New York, 2009.

[183] A. Schieben, M. Wilbrink, C. Kettwich, R. Madigan, T. Louw, and N. Merat. Designing the interaction of automated vehicles with other traffic participants: A design framework based on human needs and expectations. *Cognition, Technology and Work*, 2018.

[184] H. Schmidt, J. Terwilliger, D. AlAdawy, and L. Fridman. Hacking nonverbal communication between pedestrians and vehicles in virtual reality. In *Proc. of Driving Assessment Conference*, 2019.

[185] N. Schneider and D. M. Gavrila. Pedestrian path prediction with recursive Bayesian filters: A comparative study. In *Proc. of German Conference on Pattern Recognition*, pages 174–183, 2013.

[186] B. Schroeder, N. Rouphail, K. Salamati, E. Hunter, B. Phillips, L. Elefteriadou, T. Chase, Y. Zheng, V. P. Sisiopiku, S. Mamidipalli, et al. Empirically-based performance assessment & simulation of pedestrian behavior at unsignalized crossings. Technical report,

Southeastern Transportation Research, Innovation, Development and Education, 2014.

[187] A. T. Schulz and R. Stiefelhagen. A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction. In *Proc. of IEEE ITSC*, pages 173–178, 2015.

[188] A. T. Schulz and R. Stiefelhagen. Pedestrian intention recognition using latent-dynamic conditional random fields. In *Proc. of IEEE IV*, pages 622–627, 2015.

[189] D. C. Schwebel, D. Stavrinos, K. W. Byington, T. Davis, E. E. ONeal, and D. de Jong. Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis & Prevention*, 45:266 – 271, 2012.

[190] D. Shi, W. Zhang, and B. Wang. Modeling pedestrian evacuation by means of game theory. *Journal of Statistical Mechanics: Theory and Experiment*, (4), 2017.

[191] X. Shi, X. Shao, Z. Guo, G. Wu, H. Zhang, and R. Shibasaki. Pedestrian trajectory prediction in extremely crowded scenarios. *Sensors*, 19(5):1223, 2019.

[192] J. J. J. Slob. State-of-the-art driving simulators, a literature survey. Technical report, 2008.

[193] D. Sun, S. Ukkusuri, R. Benekohal, and S. Travis Waller. Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks. 51, 2003.

[194] L. Sun, Z. Yan, S. M. Mellado, M. Hanheide, and T. Duckett. 3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data. In *Proc. of IEEE ICRA*, 2018.

[195] L. Takayama, D. Dooley, and W. Ju. Expressing thought: improving robot readability with animation principles. *Proc. of ACM/IEEE HRI*, 2011.

[196] Y. Tamura, P. D. Le, K. Hitomi, N. P. Chandrasiri, T. Bando, A. Yamashita, and H. Asama. Development of pedestrian behavior model taking account of intention. In *Proc. of IEEE/RSJ IROS*, pages 382–387, 2012.

[197] K. Teknomo. Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model. *arXiv preprint arXiv:1610.00029*, 2016. PhD Thesis.

[198] L. L. Thompson, F. P. Rivara, R. C. Ayyagari, and B. E. Ebel. Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. *Injury Prevention*, 19(4):232–237, 2013.

[199] H. Timmermans. *Pedestrian behavior: models, data collection and applications*. Emerald, 2009.

[200] I. Tournier, A. Dommes, and V. Cavallo. Review of safety and mobility issues among older pedestrians. *Accident Analysis & Prevention*, 91:24–35, 2016.

[201] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *Proc. of IEEE/RSJ IROS*, pages 797–803, 2010.

[202] A. Turner and A. Penn. Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment. 29:473–490, 2002.

[203] A. Turnwald, D. Althoff, D. Wollherr, and M. Buss. Understanding human avoidance behavior: Interaction-aware decision making based on game theory. *International Journal of Social Robotics*, 8(2):331–351, 2016.

[204] A. Turnwald, W. Olszowy, D. Wollherr, and M. Buss. Interactive navigation of humans from a game theoretic perspective. In *Proc. of IEEE/RSJ IROS*, pages 703–708, 2014.

[205] I. University of Iowa. Driving simulator, 2018. https://www.nads-sc.uiowa.edu/.

[206] S. Vascon, E. Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, and V. Murino. A game-theoretic probabilistic approach for detecting conversational groups. In *Proc. of the Asian Conference on Computer Vision*, pages 658–675, 2014.

[207] P. Vasishta, D. Vaufreydaz, and A. Spalanzani. Natural Vision Based Method for Predicting Pedestrian Behaviour in Urban Environments. In *Proc. of IEEE ITSC*, 2017.

[208] D. Vasquez. Novel planning-based algorithms for human motion prediction. In *Proc. of IEEE ICRA*, pages 3317–3322, 2016.

[209] H. Vermuyten, J. Belin, L. D. Boeck, G. Reniers, and T. Wauters. A review of optimisation models for pedestrian evacuation and design problems. *Safety Science*, 87:167 – 178, 2016.

[210] G. Vizzari and S. Bandini. Studying pedestrian and crowd dynamics through integrated analysis and synthesis. *IEEE Intelligent Systems*, 28(5):56–60, 2013.

[211] B. Völz, H. Mielenz, G. Agamennoni, and R. Siegwart. Feature relevance estimation for learning pedestrian behavior at crosswalks. In *Proc. of IEEE ITSC*, pages 854 – 860, 2015.

[212] B. Völz, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto. Inferring pedestrian motions at urban crosswalks. *IEEE Transactions on Intelligent Transportation Systems*, 20(2):544–555, 2018.

[213] E. J. Walker, S. N. Lanthier, E. F. Risko, and A. Kingstone. The effects of personal music devices on pedestrian behaviour. *Safety Science*, 50(1):123 – 128, 2012.

[214] P. Wang, X. Huang, X. Cheng, D. Zhou, Q. Geng, and R. Yang. The apolloscape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[215] O. M. Watson and T. D. Graves. Quantitative research in proxemic behavior. *American Anthropologist*, 68(4):971–985, 1966.

[216] A. Werner. New colours for autonomous driving: An evaluation of chromaticities for the external lighting equipment of autonomous vehicles. *Colour Turn*, (1), 2019.

[217] D. Wilson, D. Wilson, and G. Grayson. *Age-related Differences in the Road Crossing Behaviour of Adult Pedestrians*. Road User Characteristics Division, Safety Department, Transport and Road Research Laboratory, 1980.

[218] C. Wojek, S. Walk, and B. Schiele. Multi-cue onboard pedestrian detection. In *Proc. of IEEE CVPR*, pages 794–801, 2009.

[219] J. Wu, J. Ruenz, and M. Althoff. Probabilistic map-based pedestrian motion prediction taking traffic participants into consideration. In *Proc. of IEEE IV*, pages 1285–1292, 2018.

[220] H. Xu, Y. Gao, F. Yu, and T. Darrell. End-to-end learning of driving models from large-scale video datasets. In *Proc. of IEEE CVPR*, pages 3530–3538, 2017.

[221] D. Yang, L. Li, K. Redmill, and Ü. Özgüner. Top-view trajectories: A pedestrian dataset of vehicle-crowd interaction from controlled experiments and crowded campus. In *Proc. of IEEE IV*, 2019.

[222] S. Yi, H. Li, and X. Wang. Pedestrian behavior understanding and prediction with deep neural networks. In *Proc. of ECCV*, pages 263–279, 2016.

[223] Y. Yuan, B. Goi-Ros, H. H. Bui, W. Daamen, H. L. Vu, and S. P. Hoogendoorn. Macroscopic pedestrian flow simulation using smoothed particle hydrodynamics (sph). *Transportation Research Part C: Emerging Technologies*, 111:334 – 351, 2020.

[224] M. Zeedyk and L. Kelly. Behavioural observations of adult-child pairs at pedestrian crossings. *Accident Analysis & Prevention*, 35(5):771 – 776, 2003.

[225] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kümmerle, H. Königshof, C. Stiller, A. de La Fortelle, and M. Tomizuka. INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps. *arXiv:1910.03088 [cs, eess]*, 2019.

[226] S. Zhang, R. Benenson, and B. Schiele. Citypersons: A diverse dataset for pedestrian detection. In *Proc. of IEEE CVPR*, pages 4457–4465, 2017.

[227] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *Proc. of IEEE ICCV*, 2015.

[228] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proc. of IEEE ICCV*, pages 3774–3782, 2017.

[229] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Proc. of IEEE CVPR*, pages 2871–2878, 2012.

[230] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey. Maximum entropy inverse reinforcement learning. In *Proc. of AAAI*, pages 1433–1438, 2008.

[231] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa. Planning-based prediction for pedestrians. In *Proc. of IEEE/RSJ IROS*, pages 3931–3936, 2009.