Understanding and Predicting Human Behavior in Interactions with Autonomous Systems in Urban Environments: A Systematic Review, Challenges, and Opportunities

Danya Li[1], Wencan Mao[2], Francisco C. Pereira[1], Yu Xiao[2], Xiang Su[3], and Rico Krueger[1]

[1]Department of Technology, Management, and Economics, Technical University of Denmark, Denmark
[2]Department of Information and Communications Engineering, Aalto University, Finland
[3]Department of Computer Science, Norwegian University of Science and Technology, Norway

Short summary

Urban transportation is undergoing a transformative shift with the advent of autonomous systems. Consequently, understanding and predicting the behaviors of vulnerable human road users (i.e., pedestrians and cyclists), including their intentions, decisions, and movements when interacting with autonomous systems becomes crucial. Reviewing the literature spanning the years 2014 to 2023, we identified 119 articles that empirically investigate the microscopic human-autonomous systems interaction (HAI). Through a systematic analysis, our paper offers a holistic overview of progress and challenges within this field, from the initial data collection and experiment design to its modeling methodologies. Based on our findings, we present our vision of bridging experimental HAI research and the works applying computer vision techniques for real-world pedestrian behavior analysis, offering valuable insights for future research directions.

Keywords: Human-autonomous system interaction, human behavior modeling, autonomous vehicles, virtual reality, controlled experiment.

1 Introduction

The urban transportation system is witnessing the rise of autonomous systems, such as autonomous vehicles (AVs), automated buses, AV shuttles, and social robots. Under this circumstance, the interaction between road users, defined as where their behavior is influenced by a space-sharing conflict (Markkula et al., 2020), is expected to change. With this paradigm shift, ensuring the safety of vulnerable road users, including pedestrians and cyclists, becomes crucial. To bridge the current non-verbal communication between drivers and humans and establish similar types of interactions between them, autonomous systems need to accurately predict human behavior, including their intentions, decisions, and movements. Thus, understanding and predicting human-autonomous system interaction (HAI) in future urban environments is of vital importance.

Studying behavior in futuristic scenarios requires data collected from both human and autonomous systems. Several data collection methods, such as wizard-of-oz (WoZ), surveys, and virtual reality (VR), are used for experimentally studying such interactions. By conducting a vast number of HAI experiments and producing unpublished datasets with diverse settings, studies are not compared to each other, obscuring the current state-of-the-art method. This lack of clarity regarding data and experiment setups, and isolated methodologies can impede the progress toward establishing safe and efficient interactions between human and autonomous systems.

Several articles give a partial overview of understanding pedestrian behavior in the presence of AVs. Rasouli & Tsotsos (2020) compare pedestrian decision-making in front of AVs and classical human-driven vehicles (HDVs) and identify practical applications for addressing the interaction problem. Tran et al. (2021) focus specifically on VR simulators-based pedestrian-AV interaction and reviewed its critical use cases, configured factors, and evaluation methods. Schneider & Bengler (2020) and Feng et al. (2021) examine pedestrian behavior in a broad sense. However, to the best of our knowledge, there is no review that encompasses the entire spectrum of research in the HAI field, from initial data collection and experiment design to modeling and implications.

The current paper aims to mitigate this research gap through a systematic review of an end-to-end process of microscopic human behavior analysis during interaction with autonomous systems from controlled experiments. We also seek to bridge insights from both HAI and the well-established pedestrian behavior prediction from practice. More specifically, we 1) summarize three main methods of studying HAI (wizard-of-oz, surveys, and VR) and highlight their strengths and limitations, 2) present the associated experimental design, and 3) examine modeling methods applied in explaining and predicting pedestrian behavior in the presence of AVs. Finally, drawing from the state-of-the-art in HAI, we pinpoint opportunities arising from both experimental and observational studies.

2 Methodology

We conduct a systematic review using the PRISMA Statement (Page et al., 2021). To identify the relevant studies, we queried two general databases: Web of Science and Scopus, and three subject-specific databases: ACM Digital Library, IEEE Xplore Digital Library, and ScienceDirect. We searched with the combination of five sets of keywords that outline our scope: "automated systems", "human road users", "controlled experiments", "interaction", and "behavior". Synonyms and related words are also used. We performed a search within the title, abstract, and keywords of each article in September 2023 and included publications dated between 2014 and 2023. The first search yielded a total of 539 records. After screening, 119 articles were identified for final analysis. The results of this paper are based on the completed 64 articles where the majority focuses on pedestrian-AV interaction, and the remainder shall be finished in March. Based on the systematic review, we conceive the framework shown in Fig. 1 following the conceptual research methodology.

3 Results

In what follows, we succinctly summarize the key takeaways of our systematic literature review.

Studying HAI using controlled experiments

Data in HAI: We divide data commonly used for studying HAI into four categories: 1) objective variables, including environmental, traffic, vehicular, and social variables; 2) personal variables, including demographic, health, and cultural characteristics; 3) psychological and physiological variables, including emotion, perceived safety, trust, eye tracking, and electroencephalogram; and 4) behavior variables, including intentions, movements, and trajectories. In HAI research, 1), 2), and 3) are used for inputs while 3) and 4) are outputs.

Data collection methods: Data in HAI can be collected from either real-world or virtual experiments. The WoZ method simulates an AV in the physical world by a disguised human driver, providing high ecological validity (i.e., enhanced ability to provide realistic experiences). However, setting up experiments and processing video recordings of interactions is time- and resource-consuming. The involvement of human drivers also introduces uncertainty in vehicle behaviors and poses issues with reproducibility. Furthermore, due to ethical considerations, people are asked to indicate their intentions, and how to transfer such intentions to behavior variables remains a concern. In contrast, VR experiments excel in controllability, safety, and cost-efficiency. Another advantage of VR experiments is the freedom to set up scenarios and the diversity of sample participants. Lastly, surveys and interviews, which provide large-scale information with limited depth of knowledge, are usually utilized as a supplement to collect subjective data (Maruhn et al., 2020).

VR techniques: Among all the VR techniques, cave automatic virtual environment (CAVE) and head-mounted display (HMD) are the most frequently used ones. CAVE is a display composed of large screens surrounding the person who perceives high-resolution computer-generated images, which has been proven self-explanatory and user-friendly, and no simulator sickness has been reported. However, the large screen and rear projectors are costly and space-consuming. Meanwhile, HMD has become increasingly popular due to its low-cost, portable, and highly immersive nature. Integrated with motion, eye-tracking, and physiological measurement systems, it enables more versatile data collection. However, HMD has a limited field of view both vertically and horizontally. Though actual walking is permitted, the presence of a head-mounted display may lead to unnatural movement. The user studies of the above-mentioned methods typically recruit 10-120 participants with dataset sizes ranging from 300 to 2400, smaller than real-world datasets collected by HDVs.

External validity of VR: Despite rapid advancements in VR techniques, the reality gap, which refers to the discrepancy between virtual simulation and reality, is still an open issue for practical application. While VR techniques provide a highly immersive environment that feels "real", there exist some essential and highly significant differences in the virtual world compared with real-world behavior. A possible reason for such differences is that participants are skeptical about the threat or damage exposed to the avatar (Serrano et al., 2022). Therefore, results from VR experiments must be interpreted with caution, especially in safety-critical applications.

Experimented traffic scenarios: The most frequent scenario (88%) is pedestrian street crossing on a road or intersection (either signalized or unsignalized, zebra or non-zebra). In this situation, people wait on the curb to reach the opposite side of the road, and AVs come from the perpendicular direction to pass humans. The two parties then interact to decide whether they yield or not and how they react depending on the situation. Fewer works (12%) extend their focus to shared spaces, presenting a more general perspective of interaction dynamics. Studying pedestrian-AV interaction in such environments is challenging due to the absence of a clear right-of-way definition; therefore, the pedestrian's behavior is harder to predict, drawing the need for more extensive HAI experiments. In terms of the involved agents, most works (89%) focus on a single pedestrian interacting with either a single or multiple AVs (possibly with mixed levels of automation). The consideration of multiple pedestrians or multi-modal traffic (e.g., vehicles, motorcycles, cyclists) in such an interaction emerges with the introduction of multi-user experiments (11%).

Human behavior modeling

Modeling objectives: We make a distinction between studies aiming to i) infer influencing factors of human behaviors , and ii) predict future behaviors according to past behaviors and influencing factors. The former objective takes up the majority (88%) while only a few recent works (12%) seek predictive ability at the same time.

Statistical methods: Human behavior explaining is mainly conducted via descriptive analysis, hypothesis testing, and statistical models while human behavior prediction is through more advanced behavior models and machine learning (ML) approaches. Among the reviewed papers, linear models in a broad sense are the most frequently used methods (i.e., linear regression, linear mixed models, generalized linear models, and generalized linear mixed models). In our view, their popularity can be ascribed to two main reasons: their close connection to repeated measures and their inherent model explainability. Collecting data from experiments requires each participant to conduct multiple trials under different or repeated scenarios and (generalized) linear mixed models enable the modeling of this hierarchically structured data. Through accompanied significance tests, variable effects can be concluded. Yet, insignificant factors in linear models can be worth exploring in non-linear ones. Additionally, these models usually do not provide high goodness-of-fit, creating opportunities for new methods.

Machine learning approaches and advanced behavior models: To better understand and predict complex human behavior, some works introduce machine learning blocks into interpretable statistical and computational models. Other works directly apply ML methods, such as support vector machines, multi-layer perceptron, and long short-term memory. These methods disregard the correlation within the repeated measure data and put emphasis on values provided by contextual variables that describe the interaction. Although these machine learning approaches generally provide higher prediction accuracy compared to their simpler counterparts, a more thorough performance comparison and considerations for interpretability and transferability are needed to conclude their superiority. Another model category comprises psychological models, which capture cognitive processes, decision-making, and behavioral patterns based on psychological theories. For instance, variable-drift diffusion models address the road-crossing problem by allowing evidence accumulation for decision-making over time. Without such a theoretical background, post-hoc analysis (e.g., Shapley value, feature importance analysis) is needed for ML methods to understand feature contributions, and this is not prevalent in the current HAI. Among all, only two works Kalatian & Farooq (2022), Pekkanen et al. (2022) test model generalizability on VR and/or real-world data.

Causal analysis: We recognize several ways to analyze human behavior and its causality. We observe that most of the works directly model the effects of objective and personal factors on human behaviors. Fewer works employ past behaviors and explicit contextual information to jointly model future behavior, resembling real-world applications using high-definition maps and sensor data. A limited amount of works use perception (i.e., psychological factors and perceived

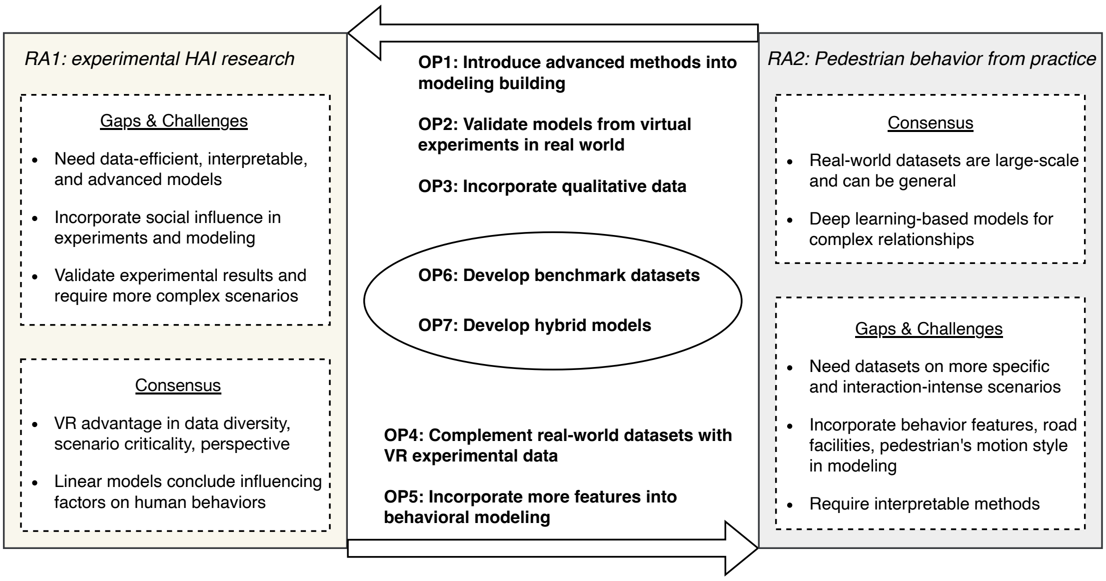

Figure 1: The envisioned framework with seven opportunities, where the circle in the middle represents shared issues that need to be addressed.

feelings) to directly model human behavior or make perception a mediator to connect observable factors and crossing behaviors.

4 Envisioned framework to combine experimental research and practical application

Motivated by the results, we propose a novel framework that combines the two highly related areas, i.e., HAI from controlled experiments (RA1) and pedestrian behavior prediction from real-world datasets (RA2). Sec. 3 concisely concludes the state-of-the-art and challenges in HAI and we adapt ideas on RA2 from Golchoubian et al. (2023); Zhang & Berger (2023). From a practical viewpoint, we consider data collection and modeling as our high-level components and provide insights on how each of these areas benefits each other, as shown in Fig. 1. The presented framework is preliminary, and it will be further substantiated upon the completion of the review of all selected studies.

In empirical research, it remains an open question whether there are significant differences in human behavior when encountering AVs and HDVs. The majority of RA1 works assume certain differences and conduct experiments of both cases accordingly. On the contrary, pedestrian behavior prediction in RA2 generally does not make such a distinction, and behavior models developed from interactions with HDVs are applied for AV development. We obey their assumptions in the following discussion.

As a natural extension of RA1, we envision incorporating RA2's modeling methods and real-world data to enhance pedestrian behavior analysis in the HAI field.

- **OP1: Introduce advanced methods into model building.** Assuming no major difference in human behavior when interacting with AVs and HDVs, models encouraging knowledge sharing are promising for future HAI research. For example, transferring pre-trained models of HDVs to AVs and analyzing the resulting difference between pre-trained and transferred models could shed light on the impacts of introducing autonomous systems. Otherwise, future HAI research requires data-efficient and interpretable models that can explain complex relationships. Models in non-supervised paradigms, meticulous model architecture design, and data augmentation techniques could help ease the sample size issue. Moreover, RA1 starts incorporating social influence from other humans in experiments, so social interaction modeling methods from pedestrian-pedestrian interaction in RA2 can help in this direction.
- **OP2: Validate predictive models from virtual experiments with real-world data.** Real-world HDV datasets capture human interaction behavior under various traffic scenarios, while virtual experiments offer specific ones. Assuming a high level of similarity in pedestrian behaviors between the two conditions, it is possible to validate the predictive

models developed from virtual environments through the real-world data generated under similar traffic scenarios. A challenge is that real-world data may not present adequate key scenarios for validation, and some readily available data in experiments may be uneasy to obtain in the real world. Going beyond validation to generalization, models developed using VR data can be tested in more general and different situations.

- **OP3: Incorporate qualitative data.** Except for accurately retrieving relationships from human movements, self-reported explanation about one's decision is another important source to utilize. Interviews of verbal clarifications are usually accompanied in experimental studies but left mainly for qualitative analysis. In the context of self-driving behavior, efforts have been made to introduce textual explainability by building large-scale datasets of descriptions and justifications related to vehicle actions Kim et al. (2018). Out of the same spirit, jointly incorporating qualitative data using natural language processing techniques and movement data can provide a new way to understand the reasoning behind pedestrian behaviors with the potential to enhance model performance.

From the other direction, we envision bringing the data diversity, scenario criticality, and multi-perspective of RA1 to address some existing challenges in RA2.

- **OP4: Complement real-world datasets with VR experimental data.** Through pre-recruitment of participants and flexible experiment setup, VR naturally provides data labeled in diverse characteristics (e.g., age, impairment, distraction, weather conditions), which can be used to test model generalizability under different conditions. Besides, many existing datasets in RA2 are captured from regular roads and few are from unregulated environments where more diverse and intensive interactions are likely to happen. VR experiments can enable a focused dataset to enrich existing general-purpose datasets. With the virtual world, developing dangerous and crash-related scenarios becomes possible. Furthermore, pedestrian datasets collected in RA2 are either from the third person or the vehicle's perspective. In contrast, VR data has rich information from the human perspective. For instance, we can observe people's gaze patterns and also obtain their physiological data. Lastly, many useful features, such as behavioral characteristics, road facilities, and pedestrian motion styles, are missing from most works, partly due to the lack of relevant datasets. Demonstrated influential factors in RA1 could guide future real-world data collection, as well as the development of camera-aided algorithms to improve data collection feasibility in the real world.
- **OP5: Incorporate more diverse features into behavioral modeling.** HAI experimental research has concluded plenty of significant factors influencing pedestrian behavior (Rasouli & Tsotsos, 2020) which can be used as model features for more accurate results. Intention prediction obtains better performance by incorporating physiological features, such as camera-based appearance and skeleton-based posture, but supplementary information about individuals (e.g., age, emotion) is not commonly used yet (Zhang & Berger, 2023). Explicitly considering these different characteristics as pedestrian styles can explore the heterogeneity within the data and allow autonomous systems to act with social intelligence.

Finally, some common issues need to be addressed before both areas can achieve a mutual benefit.

- **OP6: Develop benchmark datasets.** To study either area, there is a demand for developing benchmark datasets for both specific and general environments. This measure enables model performance evaluation and encourages progressive model development.
- **OP7: Develop hybrid models.** Requiring both model accuracy and interpretability, developing a hybrid model is expected to moderate drawbacks from either area. For example, combining existing predictive models with neuroscience-inspired ones could provide more insights into understanding and predicting human behavior.

References

Feng, Y., Duives, D., Daamen, W., & Hoogendoorn, S. (2021). Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment*, *187*, 107329.

Golchoubian, M., Ghafurian, M., Dautenhahn, K., & Azad, N. L. (2023). Pedestrian Trajectory Prediction in Pedestrian-Vehicle Mixed Environments: A Systematic Review. *IEEE Transactions on Intelligent Transportation Systems*, 1–24.

Kalatian, A., & Farooq, B. (2022). A context-aware pedestrian trajectory prediction framework for automated vehicles. *Transportation Research Part C: Emerging Technologies*, *134*, 103453.

Kim, J., Rohrbach, A., Darrell, T., Canny, J., & Akata, Z. (2018). Textual explanations for self-driving vehicles. In *Proceedings of the european conference on computer vision (eccv)* (pp. 563–578).

Markkula, G., Madigan, R., Nathanael, D., Portouli, E., Lee, Y. M., Dietrich, A., ... Merat, N. (2020). Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science*, *21*(6), 728–752.

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020). Analyzing Pedestrian Behavior in Augmented Reality — Proof of Concept. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 313–321). IEEE.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... others (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, *88*, 105906.

Pekkanen, J., Giles, O. T., Lee, Y. M., Madigan, R., Daimon, T., Merat, N., & Markkula, G. (2022). Variable-Drift Diffusion Models of Pedestrian Road-Crossing Decisions. *Computational Brain & Behavior*, *5*(1), 60–80.

Rasouli, A., & Tsotsos, J. K. (2020). Autonomous Vehicles That Interact With Pedestrians: A Survey of Theory and Practice. *IEEE Transactions on Intelligent Transportation Systems*, *21*(3), 900–918.

Schneider, S., & Bengler, K. (2020). Virtually the same? Analysing pedestrian behaviour by means of virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 231–256.

Serrano, S. M., Llorca, D. F., Daza, I. G., & Sotelo, M. A. (2022). Insertion of Real Agents Behaviors in CARLA Autonomous Driving Simulator. In *International Conference on Computer-Human Interaction Research and Applications, CHIRA - Proceedings* (Vol. 2022, pp. 23 – 31).

Tran, T. T. M., Parker, C., & Tomitsch, M. (2021). A Review of Virtual Reality Studies on Autonomous Vehicle–Pedestrian Interaction. *IEEE Transactions on Human-Machine Systems*, *51*(6), 641–652.

Zhang, C., & Berger, C. (2023). Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review. *IEEE Transactions on Intelligent Transportation Systems*, 1–23.