

A Study in Zucker: Insights on Interactions Between Humans and Small Service Robots

Alex Day  and Ioannis Karamouzas 

Abstract—Despite recent advancements in human-robot interaction (HRI), there is still limited knowledge about how humans interact and behave in the presence of small service indoor robots and, subsequently, about the human-centered behavior of such robots. This also raises concerns about the applicability of current trajectory prediction methods to indoor HRI settings as well as the accuracy of existing crowd simulation models in shared environments. To address these issues, we introduce a new HRI dataset focusing on interactions between humans and small differential drive robots running different types of controllers. Our analysis shows that anticipatory and non-anticipatory robot controllers impose similar constraints to humans' safety and efficiency. Additionally, we found that current state-of-the-art models for human trajectory prediction can adequately extend to indoor HRI settings. Finally, we show that humans respond differently to small differential drives than to other humans when collisions are imminent, since interacting with small robots can only cause a finite level of social discomfort as compared to human-human interactions.

Index Terms—Human-aware motion planning, human-robot interaction (HRI), social robot navigation.

I. INTRODUCTION

GIVEN the advancements in robotics and AI in recent years along with the significant decrease in hardware cost, robots are increasingly becoming a part of our lives, from vacuuming our floors and delivering items in automated warehouses to transporting passengers autonomously in urban areas [1]. From robot vacuums to robo-taxis, studying how humans react to such robots and interact in a shared environment plays a crucial role in the further development of autonomous systems. In this letter, we focus on small service robots like Roombas that operate in indoor spaces and have to interact intelligently with humans to avoid collisions while completing their tasks. In this field, there has been a lot of recent work

on *social robot navigation* aiming to steer robots in human-populated spaces [2]. Despite such efforts, existing approaches are typically evaluated from a robot-centric perspective. While the robot's performance is important to assess the quality of employed navigation algorithms, the efficiency and safety of the humans are equally important in shared interaction settings. Unfortunately, in such settings, there is still limited knowledge about how humans interact and behave in the presence of robots and, subsequently, about the *human-centered* behavior of robots. This also raises concerns about the applicability of current trajectory prediction methods to indoor HRI settings as well as the accuracy of existing human simulation models. To address this gap in the literature we conducted a user study in the Zucker Family Graduate Education Center in Charleston, SC, USA. Following a brief overview of highly relevant work, we focus on the related results from our analysis. Our findings and the collected HRI trajectories, which we share with the community, can enable more socially-aware robot control through providing a better understanding of the human-centered performance of different classes of robot controllers, facilitating more accurate human trajectory prediction in shared interaction settings, and informing the development of new pedestrian simulation methods.

II. RELATED WORK

A. HRI Datasets and Studies

A lot of recent work has been focusing on capturing and analyzing HRI data to better understand how humans interact with robots. The related datasets can be classified based on if the interactions they contain are unstructured or structured. Unstructured interactions can closely approximate true-to-life behavior, but require laborious manual labeling due to the lack of control in the experimental conditions [3], [4], [5]. In contrast, structured interactions occur in a controlled environment, typically equipped with sophisticated motion capture solutions that facilitate high-fidelity data collection and analysis. The THÖR dataset [6] is one such example capturing diverse indoor interactions between humans and between humans and a socially unaware small industrial robot following a predetermined path. In recent years, a number of HRI studies were conducted in structured settings to validate new methods for social navigation [7], [8], [9]. More recently, Lo et al. [10] evaluate the effect that different navigation strategies of a self-balancing robot have on humans in a perpendicular crossing scenario. Similarly, the HRI datasets in [11] and [12] study the impact that a robot has on humans during a crossing-gate and a corridor-exit experiment, respectively, with [11] showing that participants behave more conservatively in the presence of a robotic wheelchair than a

Manuscript received 14 July 2023; accepted 3 December 2023. Date of publication 18 January 2024; date of current version 1 February 2024. This letter was recommended for publication by Associate Editor D. Brscic and Editor A. Peer upon evaluation of the reviewers' comments. This work was supported by NSF under Grant IIS-2047632. (Corresponding author: Alex Day.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Clemson University under Application No. IRB2022-0251.

Alex Day is with the School of Computing, Clemson University, Clemson, SC 29634 USA (e-mail: adday@clemson.edu).

Ioannis Karamouzas is with the Clemson University, Clemson, SC 29634 USA. He is now with the Department of Computer Science and Engineering, University of California, Riverside, CA 92521 USA, and also with the CYENS - Centre of Excellence, 1016 Nicosia, Cyprus (e-mail: ioannis@cs.ucr.edu).

Our dataset and related code and models are available at: <https://motion-lab.github.io/ZuckerDataset>.

Digital Object Identifier 10.1109/LRA.2024.3355641

humanoid robot. More closely related to our work, the large-scale study of Mavrogiannis et al. [13], [14] evaluates different navigation algorithms through human-centered metrics including the energy expended by the humans and the smoothness of their paths. Our work is complimentary to such aforementioned studies as we release a new structured HRI dataset containing a range of scenarios and differential drive controllers. Similar to [10], [14] we seek to evaluate the behavior of different robot navigation methods through human-centered metrics. However, we propose an intuitive way to do so without explicitly comparing the methods but by assessing the performance of the humans with and without the robot. Unique to prior work, we provide novel results about the performance of human trajectory prediction methods in indoor HRI settings and quantify how humans resolve collisions with small service robots.

B. Social Robot Navigation Methods

There has been a lot of highly relevant work for autonomous and collision-free navigation of a robot amidst a crowd of humans [2]. Existing techniques for human-aware robot navigation can be broadly classified into geometric planning approaches and learning-based methods. State-of-the-art geometric planners rely on the concepts of velocity obstacles (VO) and time to collision [15], [16], and can provide formal guarantees of the collision-free behavior of robots while supporting a wide range of robot models and behaviors [17], [18]. However, they often require careful parameter tuning to achieve desired navigation results, with VO-based controllers also exhibiting overly conservative behavior in an attempt to guarantee safe navigation [2]. Learning-based methods can provide more flexibility, with recent contributions taking advantage of deep reinforcement learning to enable end-to-end steering and train crowd-aware navigation policies [19], [20], [21]. Despite the success stories of geometric and learning-based approaches the robots do not always exhibit the level of sophistication that humans do in similar interaction scenarios. As such, researchers have been focusing on improving the control policy of a robot by accounting for interactions between all agents in the scene, expert demonstrations, social norms, human collaboration, and group-aware planning among others [7], [9], [22], [23], [24], [25], [26], [27], [28]. In synergy with local navigation, research on HRI has also been focused on global planners to drive a mobile robot in a socially acceptable manner and facilitate human-like movements [8], [29], [30]. In this letter, we are interested in assessing the performance of three representative robot navigation algorithms in indoor physical settings. While a lot of existing works in benchmarking social robot navigation methods typically employ a robot-centric evaluation paradigm [31], [32], [33], here we follow recent work in HRI [10], [14] and focus on the *human-centered* performance of the interactions by evaluating the impact that the controlled robots have on the efficiency and safety of the humans.

C. Human Trajectory Prediction

The need for accurate methods to predict the future movements of humans is crucial to developing better robot simulation and navigation algorithms. To perform pedestrian trajectory prediction, early work has adopted model-based solutions that leverage hand-tuned mathematical models of humans' behavior. These include models based on social forces, velocity-obstacles,



Fig. 1. (left) Our tracking solution includes Vive Trackers 2018 attached to hats, a Vive Controller mounted on a Turtlebot, and Vive Base Stations 2.0. (right) A still from the overhead webcam during an experimental trial.

and data-driven formulations [15], [16], [34], [35]. In recent years, model-based methods have been replaced by model-free approaches that rely on deep learning architectures to achieve state-of-the-art (SOTA) prediction performance. Such architectures include RNN structures for generating sequential predictions, social attention and pooling mechanisms for capturing the neighbors' influence, and GAN architectures and conditional VAE (CVAE) models for accounting for the uncertainty and multimodality of human decision making [36], [37], [38], [39], [40]. We refer to the excellent survey of Rudenko et al. [41] for more details. Our paper considers two CVAE approaches that have shown to achieve SOTA performance in human-human interaction settings and evaluates their applicability to indoor HRI settings as compared to a model-based baseline.

III. EXPERIMENTAL SETUP

We conducted our experiments in lobby space near our lab measuring approximately $6\text{ m} \times 6\text{ m}$. This space represents a typical environment for the application of indoor service robots. Obstacles visible in Fig. 1 were outside the boundaries of the experimental volume. Cones marked the pedestrian start and goal positions. The participants and robot were given a simultaneous trigger to begin moving, with the participants getting an audible countdown. The Institutional Review Board of Clemson University approved the study on May 4th, 2022 under application number IRB2022-0251.

A. Participants

Two female and four male participants (age between 23 and 30 years old, $M = 26.6$, $SD = 1.6$) gave informed consent to participate in this user study. All were free of any known impediments to their walking and had normal or corrected vision, as verified by self-report. The participants were drawn from a pool of graduate students enrolled in the College of Engineering and Applied Sciences at Clemson University. The average height of the participants was $1.75 \pm 0.12\text{ m}$, and the shoulder-to-shoulder distance was $0.44 \pm 0.04\text{ m}$. None of the students had interacted with robots in their research, but most had been exposed to robotics somehow. To avoid any potential biases, the participants were not informed about the overall goal of the experiments until after the recordings had concluded. During the trials, the participants were instructed to act as they normally do when walking in our lobby space. The experiments were conducted over two consecutive days, with three participants in the first day and three in the second.

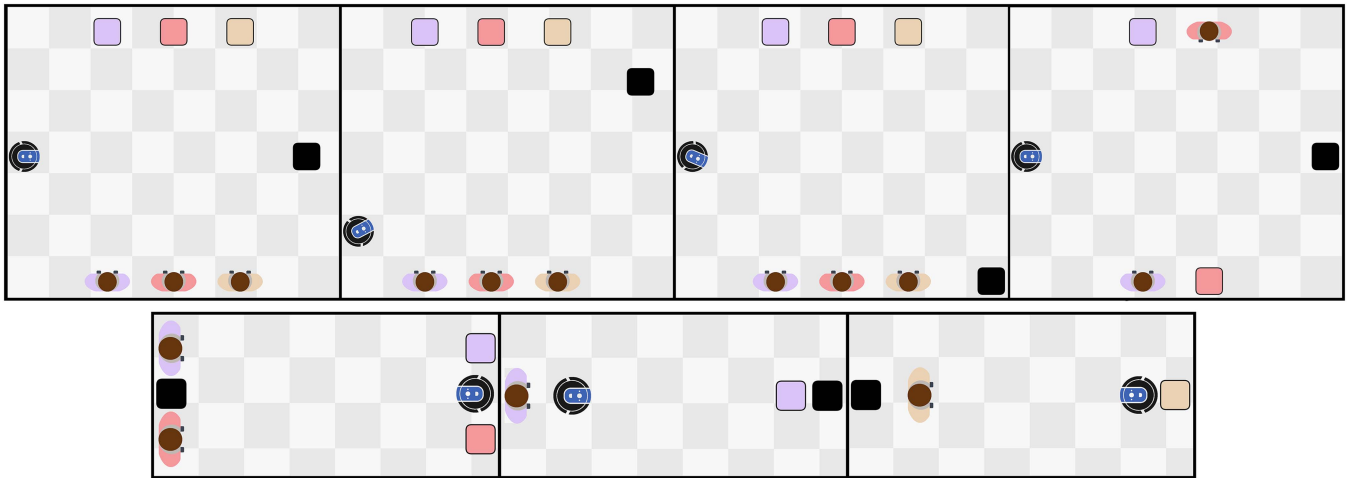


Fig. 2. Graphical representation of the interaction scenarios used in our study. The scenarios cover the entire experimental area of $6\text{ m} \times 6\text{ m}$. From top to bottom and left to right, they are *Perpendicular*, *Adjacent*, *Opposite*, *Intersection*, *2v1*, *Overtake*, and *Head-to-Head*. The top four scenarios were used in the analysis for H_1 and all seven for H_2 and H_3 . Colored squares denote the goal locations of the corresponding agents (humans or robot).

B. Robot

The wheeled robot the participants interacted with was a TurtleBot 2 from ClearPath Robotics, shown in Fig. 1. This robot is approximately 42 cm tall which may limit the applicability of our study to similarly sized robots [42]. To provide a varied dataset, we consider three local planning algorithms to steer the robot to its destination: a Linear controller that maintains a constant speed with no deviation from the straight line path to the robot's goal, NH-TTC [18] which formulates local navigation as an optimization problem employing a time-to-collision based cost function [16], and CADRL-GAC3 [21] which learns a steering policy using reinforcement learning. We chose NH-TTC and CADRL as representative geometric and machine-learning approaches for collision-free local steering, respectively. Both approaches are anticipatory in nature, accounting for potential collisions with nearby neighbors and reacting to them, as opposed to Linear, which is a non-anticipatory approach. During our experiments, all planning nodes were run on a central server and a velocity command was relayed to the robot, ensuring no delayed robot responses. The robots had a singular goal with no global plan ensuring only the local controller's behavior was analyzed. We used the official Robot Operating System (ROS) implementation of CADRL and a custom ROS wrapper over the official NH-TTC code. In all trials, we set the maximum robot speed at 0.5 m/s. The participants were not informed about the possibility of different robot controllers.

C. Interaction Scenes

We define an interaction scene as a combination of start and goal positions for participants and the robot. Fig. 2 shows the seven different scenes considered in our experiments. We refer to the our project website for example videos of the scenes. For each scene, we consider a No Robot baseline and three robot controller cases (Linear, NH-TTC, and CADRL), resulting in four different robot configurations. We define a scenario as a combination of a scene, a choice in robot configuration, and the specific order of the participants. To collect a wide array of interactions, we systematically permuted the order of the participants and the type of the robot controller to ensure that we had

a No Robot control scenario for each considered participants' order. Each participant engaged in all seven scenes traversing on average $8.8 \pm 0.2\text{ m}$ in $11.1 \pm 0.8\text{ s}$. In total, we recorded 97 robot trajectories and 309 human trajectories capturing 135 different scenarios with a total duration of 25 minutes.

D. Motion Tracking System

All trials were recorded using a tracking system based on the HTC Vive Virtual Reality system (see Fig. 1). Each participant was given a baseball cap to wear that was modified to mount an HTC Vive Tracker, allowing data collection without the use of Vive's head-mounted display. The robot was tracked with an HTC Vive Controller attached to the top shelf of the Turtlebot. The Vive System interfaced directly with a ROS node, which queried the Vive API at 60 Hz to collect the current pose of all devices. The robot's and humans' data were obtained via the use of two HTC Lighthouses positioned on either side of the recording area. The accuracy of this system was measured at the sub-centimeter level over a distance of several meters. To reduce oscillations, the recorded robot's and humans' trajectories were resampled at 60 Hz, and a Butterworth lowpass filter (2nd order, 0.5 Hz cutoff frequency) was applied to the positions. The velocity was estimated from the displacement in the smoothed positions.

IV. H_1 : ANALYSIS OF ROBOTIC CONTROLLERS

As new local planning approaches for mobile robot navigation are being introduced, it is paramount to be able to assess the performance of such approaches. While robot-centric metrics are appropriate in homogeneous environments, in shared environments, the robots should aim to maximize human safety and efficiency even at the cost of their own efficiency. As such, we consider below several human-centric metrics and propose to evaluate the robot controllers by comparing them to the No Robot case. We limit our analysis to the four scenes shown in the top row of Fig. 2. These scenes have sufficient samples to run statistical analysis and involve complex interactions containing all three participants in the environment and/or human-robot

interactions at different angles. Given their anticipatory nature, we expect that CADRL and NH-TTC robots will facilitate more optimal behaviors for all humans in the scene. Thus, we hypothesize that:

H1: CADRL and NH-TTC outperform the Linear controller in metrics related to human safety and efficiency when compared against the No Robot case.

A. Metrics

Motivated by prior work [6], [11], [13], [31], we consider three *efficiency* metrics and a *safety* metric to determine the impact each robot controller has on the participants. These metrics are calculated from the participant's perspective, rewarding human-centered behavior. Formally, given a scenario, let A denote its set of agents that include the set of human participants H and possibly a robot R . Each human participant H_i has a start position S_i , a goal position G_i , and a radius r_i determined by the shoulder-shoulder distance. Each participant also has an array of valid time steps T and, for each $t \in T$, the position, and velocity of the participant are denoted by \mathbf{p}_i^t and \mathbf{v}_i^t , respectively. The robot parameters are defined similarly. We define the following metrics on the human agents:

1) *Speed*: The speed of each human agent is calculated as the median of the L2 norm for each recorded velocity along the agent's trajectory. This metric was used to judge how unimpeded the participants movement was. The difference from their free walking speed measured at 1.3 ± 0.1 m/s before our experiments took place should also be taken into account.

2) *Travel Time*: The total time a human agent spent traveling between the start and goal positions. This metric was used to judge the overall efficiency of the human paths.

3) *Path Linearity*: The path linearity of each human agent is the sum of the absolute deviations from the straight line path from the agent's start to the goal: $\sum_{t \in T} \left| \frac{\|(G_i - S_i) \times (S_i - \mathbf{p}_i^t)\|}{\|G_i - S_i\|} \right|$. The intention of this metric is not to promote straight paths but rather to ensure that the robot doesn't impact the shape of the participant's path too drastically.

4) *Safety*: Let τ denote the time that it takes for the disks of two agents to collide assuming they both maintain their current velocities. Given that humans have a finite reaction time of 200–400 ms and an interaction time horizon of 2–4 s [16], [43], we consider any interaction with a τ -value below 1.5 s as potentially dangerous. The safety of each human agent is then defined as the ratio of the safe frames to all recorded trajectory frames: $\frac{|\{\tau(H_i^t, A_j^t) > 1.5, \forall t \in T, \forall j \neq i \in A\}|}{|T|}$, where A_j denotes either another human participant or the robot, and $\tau(H_i, A_j)$ is a function of the relative velocity, relative displacement, and radii of the two interacting agents. The intention of this metric is to highlight potentially uncomfortable interaction frames that lead to imminent collisions. We assume that controllers which minimize this metric will impede the movements of the humans.

B. Statistics

To evaluate the effect that the robot has on the participants in each of the four scenes, one-way analysis of variance was performed on each of the proposed measures using the robot type (No Robot, Linear, NH-TTC, CADRL) as independent variable. To guarantee valid comparisons, we performed a Welch ANOVA whenever the assumption of equal variance was violated and

the Kruskal-Wallis nonparametric alternative to the classic ANOVA when the normality assumption did not hold [44]. Posthoc analysis revealed no significant differences between the robot controllers in any of the studied metrics and scenes. To further investigate the human-centered behavior of the robots and determine whether introducing a robot with a specific type of controller (Linear, NH-TTC, CADRL) had a significant impact on the participants, pairwise post-hoc comparisons were performed for each scene-metric combination between each controller and the No Robot case using the latter as the control group. These were done using Dunn's test after nonparametric ANOVAs, and Dunnett's and Dunnett's T3 tests after classic and Welch ANOVAs, respectively [44].

C. Results

Regarding the safety on the *Intersection* scene, no statistical comparisons were performed given the lack of variance and large mean indicating minimal unsafe frames. For the rest of the scenes, a significant interaction effect was observed for all metrics, besides the speed on the *Adjacent* and path linearity on both *Adjacent* and *Opposite*. The mean and standard deviation for the metrics as well as the pairwise significance for the post-hoc tests are shown in Table I. Overall, when focusing on the human navigation efficiency, none of the controllers is comparable to the No Robot case for the *Opposite* scene, only the Linear controller for *Intersection*, Linear and CADRL for *Adjacent*, and the two anticipatory controllers for *Perpendicular*. Similarly, in terms of the participants' safety, no controller is universally similar to the No Robot case across all scenes. Therefore, H_1 is not supported.

D. Discussion

Our analysis focuses on the impact that a robot controller has on the navigation efficiency and safety of humans. While metrics related to speed and collisions have been used in related work, we use them in a novel way by comparing the performance of the humans with and without the robot in the scene. This provides a more intuitive way to assess the human-centered behavior of different controllers and offers an implicit comparison between them. We also explicitly compared the controllers but observed no significant pairwise differences. However, we argue that even if two controllers are statistically different from each other, if there is no statistical difference from the No Robot case, both controllers should be treated as equally unacceptable regarding human-centeredness.

In that vein, it is clear from Table I that anticipatory controllers can impede the safety and navigation efficiency of humans and do not necessarily outperform the non-anticipatory Linear controller. In retrospect, these results can be explained by the *duality* of the HRI problem. In HRI settings, the actions of a robot affect the behavior of the humans as much as the actions of the humans affect the behavior of the robot [45], [46]. So, when an NH-TTC or a CADRL Turtlebot interacts with humans, it tries to predict what the humans will do and react accordingly, sharing the collision avoidance effort. However, such intelligent, anticipatory behavior can sometimes confuse humans and negatively influence their efficiency and safety as they try to guess the robot's behavior. In contrast, when interacting with a Turtlebot that follows a straight path, the humans realize that the robot does not react to them and can quickly make minor adjustments to their trajectories way in advance to resolve

TABLE I
MEAN AND STANDARD DEVIATION FOR METRICS GROUPED BY SCENE AND CONTROLLER

Scene	Robot Case	Speed (m/s)	Travel Time (s)	Path Linearity (m)	Safety (%)
Perpendicular	Linear	0.872 ± 0.384	12.121 ± 2.213	37.737 ± 14.259	0.901 ± 0.1
	CADRL	1.114 ± 0.094	10.868 ± 1.207	31.219 ± 16.689	0.901 ± 0.11
	NHTTC	1.085 ± 0.178	10.974 ± 1.69	32.579 ± 16.264	0.866 ± 0.141
Adjacent	Linear	1.082 ± 0.156	11.417 ± 1.529	38.358 ± 29.524	0.886 ± 0.117
	CADRL	1.064 ± 0.169	11.229 ± 1.697	33.946 ± 22.505	0.878 ± 0.139
	NHTTC	0.942 ± 0.312	11.786 ± 1.307	35.605 ± 26.534	0.914 ± 0.112
Opposite	Linear	1.053 ± 0.073	11.383 ± 0.875	38.885 ± 29.95	0.925 ± 0.087
	CADRL	0.993 ± 0.171	12.252 ± 1.252	38.108 ± 27.28	0.877 ± 0.105
	NHTTC	1.012 ± 0.117	11.884 ± 1.2	34.319 ± 28.685	0.908 ± 0.094
	No Robot	1.134 ± 0.081	10.531 ± 0.509	21.431 ± 8.813	0.992 ± 0.024
Intersection	Linear	1.079 ± 0.148	10.808 ± 1.756	45.927 ± 32.669	0.997 ± 0.011
	CADRL	1.038 ± 0.094	11.1 ± 0.922	39.958 ± 22.766	1.0 ± 0.0
	NHTTC	0.999 ± 0.159	11.648 ± 1.621	47.102 ± 29.406	1.0 ± 0.0
	No Robot	1.164 ± 0.054	9.535 ± 0.646	20.29 ± 9.381	1.0 ± 0.0

Significant differences with respect to the no robot case are highlighted in red ($p < 0.05$). The no robot case is included in bold. As its statistics are identical for the 3-agent scenes in the top three rows, we only report them for the opposite.

collisions efficiently. We note that this may not be true when humans interact in more complex settings where their goals are not in direct sight. However, in our indoor settings, our analysis highlights the need for robot navigation methods that put human safety and efficiency at the forefront along with the importance of evaluating such methods from a human-centric perspective.

V. H_2 : HUMAN TRAJECTORY PREDICTION

In recent years, with the rise of deep learning techniques, many trajectory prediction approaches have shown impressive results on crowd and vehicle datasets, attributed in part to successfully reasoning about agent-agent and agent-environment interactions [41]. While such approaches can find applications to social robot navigation and facilitate more accurate human trajectory prediction, many have not been tested against human-robot-interaction scenarios [2]. To address this issue, we analyze the performance of representative model-free and model-based approaches for human trajectory prediction. Given short-term historical observations, our goal is to predict the future (ground truth) positions of the participants in our recorded interaction scenarios. A straightforward prediction approach is a constant velocity model (CVM), where the future positions of an agent are inferred from its last position and velocity, assuming a linear extrapolation. In benchmarks commonly used for human trajectory prediction, such as ETH [35], UCY [47], and SDD [48], it has been shown that CVM-based approaches can rival even state-of-the-art neural network approaches [49]. As such, we use CVM as a baseline and compare its performance to two SOTA approaches for model-free trajectory prediction, Trajectron++ [40] and SocialVAE [39]. In contrast to the deterministic nature of CVM, both Trajectron++ and SocialVAE are stochastic models outputting a future trajectory distribution by exploiting a conditional variational autoencoder architecture. Given the limited amount of complex interactions that humans undergo in our indoor interaction settings, we hypothesize that:

H2: The Constant Velocity Model outperforms SocialVAE and Trajectron++ in terms of prediction error on the Zucker dataset.

A. Evaluation Dataset and Metrics

To test H_2 , we consider all seven scenes of our Zucker dataset. Following the literature [36], [39], [40], we resample the recorded robot and human trajectories at 2.5 FPS. We further split the trajectories into segments of $H + T$ frames assuming we have access to T consecutive frames of prior observation. The problem then is to predict the future positions of each human for the next H -frames given the corresponding T -frame observation of the human's trajectory, where each observation sample includes both the position of the human and the positions of its neighbors. Given the length of the recorded participants' trajectories, we use $T = 5$ frames (2 s) as the observation window and $H = 8$ frames (3.2 s) as the prediction horizon. In total, we generated three sets of human trajectory prediction data based on the type of the robot controller that the humans interacted with. The combined dataset consists of 42,159 local trajectories that capture a variety of interactions. While the publicly available Trajectron++ and SocialVAE code offers many pre-trained models, we retrain both models to fit our observation horizons. Since the Zucker dataset does not have enough trajectories to facilitate a training-testing split, we retrained the models using the Univ dataset from the UCY benchmark [47]. This dataset contains a rich set of agents interacting at varied speeds and angles, including non-reactive and reactive agents, groups, and individuals, sufficiently covering the types of interaction scenarios that humans encountered in our experiments. As in prior work [38], [39], we use the minimum Average Displacement Error (ADE) and the minimum Final Displacement Error (FDE) over k predictions to assess a model's performance, where $k = 1$ for the deterministic CVM and $k = 5$ for the multimodal Trajectron++ and SocialVAE models. ADE measures the average L_2 distance in meters between the predicted positions and the ground truth ones over the whole predicted trajectory while FDE only focuses on the distance at the end of the predicted trajectory.

B. Results

In Table II, we report the performance of CVM, Trajectron++, and SocialVAE on the three robot scenes of our Zucker evaluation dataset. Overall, the performance of the CVM baseline can

TABLE II
ADE/FDE IN METERS OF CVM, TRAJECTRON++, AND SOCIALVAE IN ZUCKER
SCENES GROUPED BY ROBOT CONTROLLERS

Method	CADRL	NH-TTC	Linear	Average
CVM	0.21 / 0.42	0.19 / 0.38	0.20 / 0.41	0.20 / 0.40
Trajectron++	0.17 / 0.33	0.16 / 0.32	0.14 / 0.28	0.16 / 0.31
SocialVAE	0.16 / 0.32	0.15 / 0.29	0.15 / 0.30	0.15 / 0.30

We report the best of 5 predictions for Trajectron++ and SocialVAE.

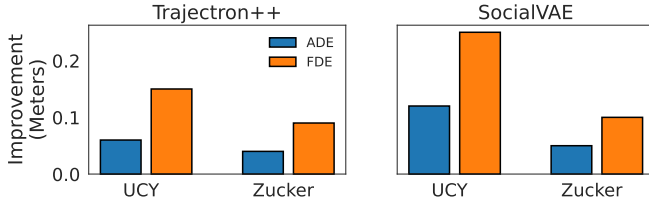


Fig. 3. ADE/FDE from Trajectron++ and SocialVAE on the UCY and Zucker datasets in terms of improvement over the CVM. The Zucker results are obtained using all three robot steering controllers.

provide a quick way to assess the interaction complexity in the scene [49]. Small ADE/FDE values typically denote that humans exhibit mostly linear interactions, with the past trajectory of a tracked human being the main feature for predicting its future. Given the CVM performance, one could argue that humans do not undergo too complex social interactions in Zucker. Therefore, employing a linear model to track the trajectories of humans is a strong baseline. However, as seen in Table II, the model-free approaches attain lower prediction errors, outperforming the CVM baseline across the different scenes. As such, we reject H_2 .

C. Discussion

Our analysis shows that model-free approaches such as Trajectron++ and SocialVAE can provide reliable human trajectory predictions in small indoor settings that can potentially inform more human-aware robot control. On the other hand, a linear model can serve as a strong baseline in similar HRI indoor settings, especially as more complex methods might introduce unnecessary computational overhead in an already strained real-time setting. To put the above results into perspective, we also compute the performance of the three studied trajectory prediction approaches on the pedestrian datasets from the UCY benchmark. In Fig. 3, we report the performance improvement of Trajectron++ and SocialVAE over CVM for both the UCY and Zucker datasets. Similar to Zucker, the model-free approaches outperform CVM. However, the performance improvement is much lower in our Zucker dataset than in the human-only UCY datasets commonly used in trajectory prediction benchmarks. This indicates that there is potential room to improve the performance of human trajectory prediction methods in indoor HRI settings by, e.g., training models with more representative datasets, devising new models, and/or refining trained models both in online and lifelong manners [50].

VI. H_3 : HUMAN-ROBOT INTERACTIONS

In [16] it was shown that the “interaction energy” between pairs of humans in a crowd can be modeled through an inverse power-law relationship with their projected time to collision, τ .

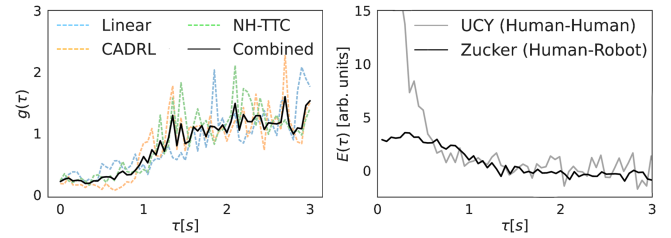


Fig. 4. (left) The pair distribution function g as a function of the projected time-to-collision value τ between human-robot pairs. The curves are grouped based on the type of controller that the robot employs and exhibit similar behavior, allowing us to combine them. (right) The corresponding interaction energy E remains finite even for small τ -values, as opposed to the human-human interaction energy inferred from the UCY crowd dataset (obtained from [16]). The energy curves are normalized so that $E(1) = 1$.

The intuition is that the sooner two pedestrians are going to collide, the more uncomfortable they feel leading to a high energy to avert the collision. This mathematical model of human-human interactions is therefore anticipatory, depending not just on the current state of the environment but also on expected future states. While its applicability has been validated in many different conditions and settings, it was derived from human-only datasets. Here we are interested in gaining a better understanding of heterogeneous interactions between humans and small service robots in indoor spaces like the ones considered in our user study. We believe that, in such settings, humans will be more cautious and try to avoid collisions with the robot as early as possible. As such, we hypothesize that:

- H3: The interaction energy between human-robot pairs in the Zucker dataset is different than that observed between human-human pairs in crowd-only datasets.

A. Methodology

Following [16], we focus on the time to collision metric, τ , and employ a probabilistic analysis approach to study pairwise interactions in isolation from all other factors that can influence the behavior of agents. To do so, we compare the frequency of τ -values between human-robot pairs that appear in a recorded scene at the same time to the frequency of τ -values for random, non-interacting pairs. The resulting normalized frequency estimates the pair distribution function, $g(\tau)$, that highlights statistically suppressed interactions. Such a function can be converted to a “social interaction energy”, E , assuming a Boltzmann-like relation between g and E : $E(\tau) \propto \log(1/g(\tau))$. Thus, the energy vanishes when $g(\tau) = 1$ and drastically increases for small $g(\tau)$ values.

B. Results

Fig. 4 left shows the pair distribution function $g(\tau)$ from the Zucker dataset. The reported plots were obtained by considering τ -values for human-robot-only pairs. Overall, the NH-TTC, CADRL, and Linear curves show very similar behavior, allowing us to combine them and estimate the normalized distribution of the τ -values between the human participants and robot controllers. We note that during our HRI experiments, no actual collisions took place, though several near collisions were observed (the positive $g(\tau)$ value for $\tau = 0$ is due to the binning of τ). In Fig. 4 right, we show the interaction energy $E(\tau)$ derived from $g(\tau)$. We also depict the interaction energy graph

from [16] obtained by analyzing pedestrian interactions in the UCY crowd scenes. As can be seen, the Zucker and UCY energy plots are visibly different. A two-sample Kolmogorov-Smirnov test statistically confirmed that the respective $E(\tau)$ values are coming from different distributions. Thus, we accept H_3 .

C. Discussion

Our results show that in both our HRI scenes and the crowd interaction scenes, humans don't care about collisions that will take place in the far future ($E(\tau) \approx 0$ for $\tau > 1.5$ s in both UCY and Zucker). Furthermore, humans anticipate collisions when interacting with other agents (humans or Turtlebots), with the energy increasing with decreasing time-to-collision values. However, when faced with an imminent collision, and in contrast to what we expected before the experiments, the interaction energy between a human and a Turtlebot is finite as opposed to human-human interactions that exhibit infinite energy. Specifically, $E(\tau)$ rises to infinity when $\tau < 0.3$ for the UCY pairs, while it saturates around a constant value for such τ 's in Zucker. In hindsight, such a finding is not that surprising. Turtlebots can only cause a finite amount of social discomfort, as they move slowly and are not intimidating. For example, we typically step over Roombas, highlighting our lack of any fear of colliding with them. Hence, humans are more willing to head toward a collision since they can resolve it quickly. We expect the observed interactions to extend to other small indoor robots leading to the development of more accurate models of human behavior compared to the ones currently employed in robot simulators [15], [16], [33], [34]. We acknowledge, though, that such finite interaction forces may not apply to larger indoor robots such as humanoids or fast-moving transport robots deployed in fulfillment centers.

VII. CONCLUSION

In this letter, we study how human behavior is affected by different types of robot controllers in indoor navigation settings. We conclude by highlighting some of the limitations and potential applications of our user study.

Limitations: Our work focuses on sparse interactions between humans and small differential drive robots, and thus the conclusions drawn from our study do not necessarily translate to other settings or types of robots. Although, we believe that the interactions are representative of ones that humans could experience in their households and workspaces. Our experiments took place in the lobby area of a building providing a more real-world setting than a traditional laboratory space. However, the participants were aware that they were involved in a user study, which could influence their behavior. Finally, we note that our user study includes a small number of participants which can limit the generalization of our results. Still, our analysis can be broadened with further data collection, an avenue we aim to explore in subsequent studies.

Applications and Future Work: Despite the aforementioned limitations, we believe that our work can facilitate future research in *social robot navigation*, *human trajectory prediction*, and *crowd simulation*. In particular, regarding *social robot navigation*, we propose an intuitive way to measure the *human-centered* behavior of a robot by comparing the performance of humans with and without the robot. The related analysis shows the need to invest in more human-centered methods for

mobile indoor navigation [51], seeking to prioritize human over robot performance (e.g., for tasks like vacuuming that aren't constrained by time). Our evaluation approach could also be applied to recent work that has explored group-aware robot planning techniques and collaborative strategies for robot collision avoidance, among others [24], [25], [26], [52].

Regarding *human trajectory prediction*, the performance results reported in Table II can serve as a strong baseline for testing other prediction methods. We further release our evaluation HRI dataset along with our pretrained models to facilitate the refinement of existing trajectory prediction methods and the development of new ones [41]. Similarly, our dataset can find applications to continual learning methods that can help improve existing trajectory prediction models in settings with continuously adapting human behavior [50]. Finally, our dataset can inform the development of more accurate *crowd simulation* models that can facilitate the training and evaluation of robot navigation methods. In existing robot simulators and benchmarking tools [21], [23], [32], [33], human behavior is typically modelled with distance-based potentials [34] or anticipatory ones based on the notion of time to collision, τ [15], [16]. However, our analysis suggests a combination of the two may be a more accurate model for human-robot interaction. As such, the corresponding $E(\tau)$ plot in Fig. 4 can be used to derive a geometric model for human-robot interactions. Alternatively, our collected human trajectories can serve as expert demonstrations to train data-driven crowd models through reinforcement learning [7], [9].

REFERENCES

- [1] L. Eliot, "Self-driving cars and Asimov's three laws about robots," *Forbes*, Jan. 2021. [Online]. Available: <https://www.forbes.com/sites/lanceeliot/2021/01/05/self-driving-cars-and-asimovs-three-laws-about-robots/?sh=140036b94768>
- [2] C. Mavrogiannis et al., "Core challenges of social robot navigation: A survey," *ACM Trans. Hum.-Robot. Interact.*, vol. 12, no. 3, pp. 1–39, 2023.
- [3] Z. Yan, S. Schreiberhuber, G. Halmetschlager, T. Duckett, M. Vincze, and N. Bellotto, "Robot perception of static and dynamic objects with an autonomous floor scrubber," *Intell. Serv. Robot.*, vol. 13, no. 3, pp. 403–417, 2020.
- [4] R. Martin-Martin et al., "JRDB: A dataset and benchmark of egocentric robot visual perception of humans in built environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 6748–6765, Jun. 2023.
- [5] H. Karnan et al., "Socially compliant navigation dataset (SCAND): A large-scale dataset of demonstrations for social navigation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 11807–11814, Oct. 2022.
- [6] A. Rudenko, T. P. Kucner, C. S. Swaminathan, R. T. Chadalavada, K. O. Arras, and A. J. Lilienthal, "Thör: Human-robot navigation data collection and accurate motion trajectories dataset," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 676–682, Apr. 2020.
- [7] B. Kim and J. Pineau, "Socially adaptive path planning in human environments using inverse reinforcement learning," *Int. J. Soc. Robot.*, vol. 8, pp. 51–66, 2016.
- [8] X.-T. Truong and T. D. Ngo, "Toward socially aware robot navigation in dynamic and crowded environments: A proactive social motion model," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 4, pp. 1743–1760, Oct. 2017.
- [9] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, "Socially compliant mobile robot navigation via inverse reinforcement learning," *Int. J. Robot. Res.*, vol. 35, no. 11, pp. 1289–1307, 2016.
- [10] S.-Y. Lo, K. Yamane, and K.-I. Sugiyama, "Perception of pedestrian avoidance strategies of a self-balancing mobile robot," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2019, pp. 1243–1250.
- [11] B. Zhang, J. Amirian, H. Eberle, J. Pettré, C. Holloway, and T. Carlson, "From HRI to CRI: Crowd robot interaction understanding the effect of robots on crowd motion," *Int. J. Soc. Robot.*, vol. 14, no. 3, pp. 631–643, 2022.

- [12] Z. Chen, C. Jiang, and Y. Guo, "Pedestrian-robot interaction experiments in an exit corridor," in *Proc. IEEE 15th Int. Conf. Ubiquitous Robot.*, 2018, pp. 29–34.
- [13] C. Mavrogiannis, A. M. Hutchinson, J. Macdonald, P. Alves-Oliveira, and R. A. Knepper, "Effects of distinct robot navigation strategies on human behavior in a crowded environment," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2019, pp. 421–430.
- [14] C. Mavrogiannis, P. Alves-Oliveira, W. Thomason, and R. A. Knepper, "Social momentum: Design and evaluation of a framework for socially competent robot navigation," *ACM Trans. Hum.-Robot Interact.*, vol. 11, no. 2, pp. 1–37, 2022.
- [15] J. van den Berg, S. J. Guy, M. Lin, and D. Manocha, "Reciprocal n -body collision avoidance," in *Robot. Res.*, Berlin, Germany: Springer, 2011, pp. 3–19.
- [16] I. Karamouzas, B. Skinner, and S. J. Guy, "Universal power law governing pedestrian interactions," *Phys. Rev. Lett.*, vol. 113, no. 23, 2014, Art. no. 238701.
- [17] Y. Luo, P. Cai, A. Bera, D. Hsu, W. S. Lee, and D. Manocha, "PORCA: Modeling and planning for autonomous driving among many pedestrians," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3418–3425, Oct. 2018.
- [18] B. Davis, I. Karamouzas, and S. J. Guy, "NH-TTC: A gradient-based framework for generalized anticipatory collision avoidance," in *Proc. Robot. Sci. Syst.*, 2020, Paper no. 078.
- [19] T. Fan, P. Long, W. Liu, and J. Pan, "Distributed multi-robot collision avoidance via deep reinforcement learning for navigation in complex scenarios," *Intl. J. Robot. Res.*, vol. 39, no. 7, pp. 856–892, 2020.
- [20] A. J. Sathiamoorthy, J. Liang, U. Patel, T. Guan, R. Chandra, and D. Manocha, "DenseCAvoid: Real-time navigation in dense crowds using anticipatory behaviors," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2020, pp. 11345–11352.
- [21] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," *IEEE Access*, vol. 9, pp. 10357–10377, 2021.
- [22] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2017, pp. 1343–1350.
- [23] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2019, pp. 6015–6022.
- [24] C. Mavrogiannis, K. Balasubramanian, S. Poddar, A. Gandra, and S. S. Srinivasa, "Winding through: Crowd navigation via topological invariance," *IEEE Robot. Autom. Lett.*, vol. 8, no. 1, pp. 121–128, Jan. 2023.
- [25] K. Katyal et al., "Learning a group-aware policy for robot navigation," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2022, pp. 11328–11335.
- [26] D. Gonon and A. Billard, "Inverse reinforcement learning of pedestrian-robot coordination," *IEEE Robot. Autom. Lett.*, vol. 8, no. 8, Aug. 2023.
- [27] P. Bevilacqua, M. Frego, D. Fontanelli, and L. Palopoli, "Reactive planning for assistive robots," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 1276–1283, Apr. 2018.
- [28] S. Liu et al., "Intention aware robot crowd navigation with attention-based interaction graph," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2023, pp. 12015–12021.
- [29] R. Kirby, "Social robot navigation," *Ph.D. dissertation, Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA*, 2010.
- [30] T. Kruse, P. Basili, S. Glasauer, and A. Kirsch, "Legible robot navigation in the proximity of moving humans," in *Proc. IEEE Workshop Adv. Robot. Social Impacts*, 2012, pp. 83–88.
- [31] A. Biswas, A. Wang, G. Silvera, A. Steinfeld, and H. Admoni, "Soc-NavBench: A grounded simulation testing framework for evaluating social navigation," *ACM Trans. Hum.-Robot Interact.*, vol. 11, no. 3, pp. 1–24, 2022.
- [32] N. Tsoi et al., "SEAN 2.0: Formalizing and generating social situations for robot navigation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 11047–11054, Oct. 2022.
- [33] L. Kästner et al., "Arena-bench: A benchmarking suite for obstacle avoidance approaches in highly dynamic environments," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 9477–9484, Oct. 2022.
- [34] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.
- [35] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 261–268.
- [36] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2016, pp. 961–971.
- [37] A. Vemula, K. Muelling, and J. Oh, "Social attention: Modeling attention in human crowds," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 4601–4607.
- [38] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2018, pp. 2255–2264.
- [39] P. Xu, J.-B. Hayet, and I. Karamouzas, "SocialVAE: Human trajectory prediction using timewise latents," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 511–528.
- [40] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron : Dynamically-feasible trajectory forecasting with heterogeneous data," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 683–700.
- [41] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *Int. J. Rob. Res.*, vol. 39, no. 8, pp. 895–935, 2020.
- [42] I. Rae, L. Takayama, and B. Mutlu, "The influence of height in robot-mediated communication," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2013, pp. 1–8.
- [43] A.-H. Olivier, A. Marin, A. Crétual, and J. Pettré, "Minimal predicted distance: A common metric for collision avoidance during pairwise interactions between walkers," *Gait Posture*, vol. 36, no. 3, pp. 399–404, 2012.
- [44] J. Hsu, *Multiple Comparisons: Theory and Methods*. Boca Raton, FL, USA: CRC, 1996.
- [45] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state," in *Proc. IEEE Int. Conf. Intell.*, 2016, pp. 66–73.
- [46] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2010, pp. 797–803.
- [47] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 655–664, 2007.
- [48] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory prediction in crowded scenes," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 549–565.
- [49] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll, "What the constant velocity model can teach us about pedestrian motion prediction," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1696–1703, Apr. 2020.
- [50] L. Knoedler, C. Salmi, H. Zhu, B. Brito, and J. Alonso-Mora, "Improving pedestrian prediction models with self-supervised continual learning," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4781–4788, Apr. 2022.
- [51] C. Vassallo, A.-H. Olivier, P. Souères, A. Crétual, O. Stasse, and J. Pettré, "How do walkers behave when crossing the way of a mobile robot that replicates human interaction rules?," *Gait Posture*, vol. 60, pp. 188–193, 2018.
- [52] M. Sun, F. Baldini, P. Trautman, and T. Murphey, "Move beyond trajectories: Distribution space coupling for crowd navigation," in *Proc. Conf. Robot. Sci. Syst.*, 2021, Paper no. 053.