Characterizing the Complexity of Social Robot Navigation Scenarios

Andrew Stratton[1], Kris Hauser[2], and Christoforos Mavrogiannis[1]

***Abstract*—Social robot navigation algorithms are often demonstrated in overly simplified scenarios, prohibiting the extraction of practical insights about their relevance to real-world domains. Our key insight is that an understanding of the inherent complexity of a social robot navigation scenario could help characterize the limitations of existing navigation algorithms and provide actionable directions for improvement. Through an exploration of recent literature, we identify a series of factors contributing to the complexity of a scenario, disambiguating between contextual and robot-related ones. We then conduct a simulation study investigating how manipulations of contextual factors impact the performance of a variety of navigation algorithms. We find that dense and narrow environments correlate most strongly with performance drops, while the heterogeneity of agent policies and directionality of interactions have a less pronounced effect. Our findings motivate a shift towards developing and testing algorithms under higher-complexity settings.**

***Index Terms*—Human-Aware Motion Planning, Human-Centered Robotics, Autonomous Vehicle Navigation.**

I. Introduction

RECENT surveys on social robot navigation [1–3] have highlighted that strong assumptions in typical evaluation practices prohibit transfer to real-world domains. Much of the literature considers sparse, slowly navigating human crowds, giving rise to non-interactive human-robot encounters that are easily handled by classical navigation algorithms. Hence, the benefits of many modern approaches are not transparent. This underscores the need for standardized benchmarks; however, defining benchmarks for this domain requires community consensus over appropriate experimental design. This would need to capture several aspects, such as the spatial arrangement and roles of agents, the briefing that human users receive before the experiment, robot design specifications, metrics evaluating robot performance and human impressions, etc.

In this paper we argue that to establish effective benchmarks for social robot navigation, the research community should better understand and control the *dimensions of problem Complexity*. Clear definitions of Complexity in fields like theoretical computer science [4], learning theory [5], and motion planning [6] enable a characterization of limitations in existing algorithms and motivate future research directions. However, robotics is a synthetic science [7], combining many elements from other fields, which makes it harder to formulate similarly succinct and relevant definitions for Complexity. Social robot navigation [1] is no exception to this complication, as it lies at the intersection of multiple research areas, including motion planning, machine learning, control, and human modeling, among others. In this work, we approach Complexity through an empirical lens: a dimension of *Complexity* is a factor that causes a drop in *Performance* of an automated method (Fig. 1). Here, Performance is measured with respect to relevant criteria accepted by the community [1].

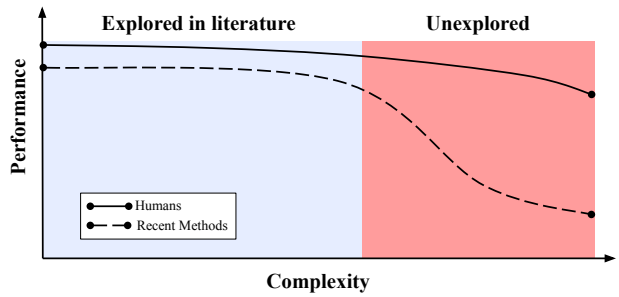


Fig. 1: Humans can seamlessly handle a wide range of crowd navigation scenarios [8]. In contrast, social robot navigation algorithms struggle to handle scenarios with realistic levels of Complexity [9]. In this paper, we show that high-Complexity scenarios are underexplored in the literature. We demonstrate principal factors contributing to poor navigation performance, and argue for the definition of benchmarks that account for their impact.

To instantiate our framework, we review the evaluation practices of recent literature and extract a set of factors contributing to the Complexity of a social robot navigation scenario, distinguishing between contextual and robot-related factors. We then conduct an extensive simulated study to understand how robot-independent, contextual factors impact algorithmic performance. We find that Density and Environment Geometry have the strongest correlation with performance decline across a wide range of algorithms. We also find that the scenarios most frequently explored in prior work tend to be of lower Complexity, and allow simple, reactive methods to perform effectively. Our findings underscore the need for algorithms that explicitly account for diverse, high-Complexity scenarios, and benchmarks that explore performance in scenarios that have high Complexity across multiple factors.

II. Related Work

Recent surveys on social robot navigation [1–3, 10] emphasize the need for a standardization of benchmarking practices.

Manuscript received: May, 18, 2024; Revised August, 21, 2024; Accepted October, 28, 2024.

This paper was recommended for publication by Editor Gentiane Venture upon evaluation of the Associate Editor and Reviewers' comments. A. Stratton was partially funded by NSF Grant # NRI-2025782.

[1]Andrew Stratton and Christoforos Mavrogiannis are with the Department of Robotics, University of Michigan, Ann Arbor, USA. {arstr, cmavro}@umich.edu.

[2]Kris Hauser is with the Department of Computer Science, University of Illinois at Urbana-Champaign. kkhauser@illinois.edu.

Digital Object Identifier (DOI): see top of this page.

Mavrogiannis et al. [1] observe the lack of rigorous statistical characterization of robot performance and human impressions, and emphasize the need for better repeatability. Francis et al. [2] collect a list of principles that a social navigation system should take into account, and propose guidelines for evaluating them. Gao and Huang [10] delve into the evaluation methodologies adopted by researchers and practitioners. Singamaneni et al. [3] provide a taxonomy of the literature, capturing aspects such as the robot type, the algorithmic approaches used, and the social navigation context. This work complements prior surveys by analyzing the Complexity of the scenarios employed in the evaluation practices of published approaches. By understanding the explored territory of Complexity, we provide directions for integrating more Complex settings in benchmarking practices that more closely resemble what a robot would experience in the real world.

A major challenge in standardizing benchmarking in social navigation lies in the design of experiments, and specifically in balancing repeatability and emergence of natural interactions. Mavrogiannis et al. [1] classify real-world experiments into three classes: demonstrations [11–13], lab studies [14–18], and field studies [9, 19, 20], advocating for the extraction of statistical insights across social navigation experiments. However, the cost of conducting real-world experiments has motivated experimentation within virtual environments [21]. Borrowing tools from graphics and crowd simulation [22, 23], several works have developed photorealistic simulators [21, 24, 25], while others have focused on finding quantitative metrics that capture aspects of performance previously only understood through qualitative metrics [26–28]. However, as demonstrated by Fraichard and Levesy [29], simulation is often introduced in evaluation practices with shortcomings. Common assumptions in simulated evaluations include the robot being *invisible* to humans, humans being simulated as homogeneous overly submissive agents, the scenarios considered being generated at random, and the scenarios containing no walls or obstacles. Such assumptions result in surprising observations, such as that a "blind" robot moving straight can outperform state-of-the-art agents [1]. Motivated by these observations, we contribute an experimental design that gives rise to more realistic crowd conditions by considering the manipulation of contextual variables including the Environment Geometry, the agent Policy Mixture, the crowd Directionality, and Density.

Another challenge in designing benchmarks is the lack of understanding of what makes a social navigation scenario Complex. In prior work, the crowd Density [1, 9, 30] has often been used as a proxy for Complexity. However, Density on its own does not capture the Complexity of the motion coupling between closely interacting agents. To capture that, metrics like Path Irregularity [31] and the Topological Complexity Index [27, 32] quantify aspects of geometric and topological richness of agents' interactions. Our work is relevant to these approaches, however it seeks to extract a more fundamental understanding of how the parameters defining a social navigation scenario impact its Complexity. We are inspired by the Complexity definitions of theoretical fields [4–6]. While our investigation is empirical, we see it as a first step towards defining a formal representation of Complexity.

III. Factors of Complexity in the Literature

We first identified factors of Complexity in the literature by investigating benchmarking practices in papers from the ICRA, IROS, CoRL, and RSS conferences from 2015 to 2024, using the keywords "social" and "crowd". Additionally, we reviewed work cited in recent surveys [1–3, 10]. We studied a variety of evaluation methodologies and found a helpful breakdown of the factors characterizing the evaluation of each scenario could be through two categories: contextual (see Fig. 2) and robot-related factors. We tabulate the results in Table I.

A. Contextual Factors

By contextual, we refer to factors are *robot-agnostic*, i.e. they are related to the environment the robot is deployed in.

Density. Density is often used as a proxy for the Complexity of a scenario [1, 9, 30]. We report it in the standard form of $agents/m^2$, although we note this does not account for the variable size of agents. Wherever a Density was not explicitly given or computable, we approximated based on supplemental footage to complete Table I.

Directionality. The directions in which humans encounter the robot contribute to the difficulty of the collision-avoidance task [62]. We identified four cases– Passing: Agents move parallel, paths do not intersect; Crossing: Agents move perpendicular, paths intersect; Random: Agent starts and goals are randomly sampled; Circle Crossing: Agent start and goals are sampled on opposite sides of the circumference of a circle.

Environment Geometry. We find that most often, evaluations take place in either hallways or medium to large office rooms, which place no constraints on the agents' movements.

Policy Mixture. The behavior with which each agent co-navigates affects the evolution of a scenario. Cooperative agents assume partial responsibility for collision avoidance, which simplifies the robot's task. In contrast, when rushing, being distracted, or changing intentions, humans may pose greater challenges to a robot [16]. We found that most real-world studies instruct participants to navigate cooperatively, and most simulation-based studies make use of cooperative crowd simulators like Social Force [23] and ORCA [63].

B. Robot Factors

These factors refer to details of the robot hardware platform, and include the robot footprint, mass, maximum speed, and sensing capabilities, among others.

Robot Size. Most recent evaluations use robots with a smaller form factor than those in previous works, which allows other agents to maneuver around them more easily.

Maximum Speed. Consistently, the maximum speed of the robot is much lower than a natural human walking speed [16], due to hardware or safety constraints. When reviewing supplementary video material, however, we observed that this often resulted in the humans moving to avoid the robot, while the robot itself performed little to no socially aware maneuvers.

Kinematics Most often, we see differential drive robots used in real world experiments. This is seldom matched in simulation training and testing, in which it is frequently assumed the robot is holonomic [1, 36, 64].

TABLE I: Complexity factors in the literature. Directionality: Passing (P), Crossing (C), Circle Crossing (CC), Random (R); Density: Low $[0, 0.1)$, Medium $[0.1, 0.25]$, High $(0.25, \infty)$ $agents/m^2$; Environment: Small room $[0, 25)$, Medium room $[25, 100]$, Large room $(100, \infty)$ m^2.

Paper	Density	Directionality	Environment	Robot Size (mm^3)	$v_{max}(m/s)$	Kinematics	Sensors	Viewpoint
Angelopoulos et al. [33]	Low	P	Hallway	425x480x1210	0.85	DD	N/A	Egocentric
Bennewitz et al. [34]	Low	N/A	Office	520x520x1180	0.65	N/A	Stereo, LIDAR	Egocentric
Burgard et al. [20]	Medium	N/A	Museum	N/A	0.80	N/A	LIDAR	Egocentric
Ciou et al. [35]	Low	P	Hallway	381x455x217	1.20	DD	LIDAR, RGB-D	Egocentric
Chen et al. [36]	Medium	P, C	Hallway	N/A	N/A	DD	N/A	N/A
Katyal et al. [37]	Low	P	Med. Room	1100x500x610	1.60	Quad	Motion Capture	Overhead
Kim and Pineau [38]	N/A	P	Med. Room	N/A	N/A	N/A	RGB-D	Egocentric
Kirby [39]	Low	P	Hallway	520x520x1180	0.70	N/A	Laser scanner	Egocentric
Liu et al. [13]	Medium	N/A	Med. Room	354x354x420	0.50	DD	LIDAR	Egocentric
Liu et al. [40]	Medium	P, C	Outdoor	N/A	N/A	DD	LIDAR	Egocentric
Liu et al. [12]	Low	P, R	Med. Room	354x354x420	0.65	DD	RGB-D	Egocentric
Lo et al. [18]	Low	P	L. Room	N/A	1.50	Omni	Motion Capture	Overhead
Matsuzaki et al. [41]	Medium	P, C	L. Room	N/A	1.20	LIDAR	N/A	Egocentric
Matsuzaki et al. [42]	High	P, C	L. Room	N/A	0.50	N/A	N/A	N/A
Mavrogiannis et al. [15]	Medium	P, C	Sm. Room	N/A	0.75	Omni	Motion Capture	Overhead
Mavrogiannis et al. [27]	Medium	P, C	Med. Room	508x660x1580	0.83	Omni	Motion Capture	Overhead
Mun et al. [43]	Low	P	L. Room	354x354x420	0.65	DD	LIDAR	Egocentric
Mustafa et al. [44]	Low	P	Med. Room	504x430x250	2.00	DD	N/A	N/A
Nguyen et al. [45]	Low	P	Hallway	810x320x640	4.80	Ackermann	LIDAR	Egocentric
Oh et al. [46]	Low	P, C	L. Room	504x430x250	2.00	DD	LIDAR, Stereo	Egocentric
Pacchierotti et al. [17]	Low	P	Hallway	N/A	0.60	N/A	LIDAR	Egocentric
Paez-Granados et al. [47]	High	P, C	Outdoor	N/A	N/A	N/A	LIDAR, RGB-D	Egocentric
Peddi et al. [48]	Medium	P	Sm. Room	960x793x296	0.50	Omni	Motion Capture	Overhead
Poddar et al. [16]	Medium	P, C	Sm. Room	N/A	1.50	Omni	Motion Capture	Overhead
Qin et al. [49]	Low	P, C	L. Room	N/A	0.85	DD	LIDAR	Egocentric
Qiu et al. [50]	Medium	P, C	Hallway	354x354x420	0.65	DD	LIDAR, RGB-D	Egocentric
Sathyamoorthy et al. [51]	High	N/A	Hallway	354x354x420	1.00	DD	LIDAR, RGB-D	Egocentric
Shiomi et al. [52]	N/A	N/A	L. Room	600x600x1200	0.75	N/A	LIDAR	Overhead
Silva et al. [53]	Low	P	Med. Room	425x480x1210	0.55	DD	Motion Capture	Overhead
Singamaneni et al. [54]	Low	P	Doorway	668x668x1640	1.00	N/A	LIDAR	Egocentric
Tai et al. [11]	Low	P, C	M. Room	281x306x141	0.26	DD	LIDAR, RGB-D	Egocentric
Thrun et al. [19]	Medium	N/A	Museum	N/A	0.70	N/A	LIDAR	Egocentric
Trautman et al. [9]	High	P, C	Med. Room	N/A	0.30	N/A	Stereo	Overhead
Truong and Ngo [55]	Low	P	Med. Room	N/A	N/A	N/A	LIDAR, RGB-D	Egocentric
Tsai and Oh [56]	Medium	P, C	Road	990x670x390	1.00	DD	LIDAR	Egocentric
Wang et al. [57]	Medium	CC	L. Room	N/A	N/A	N/A	RGB-D	Egocentric
Wang et al. [58]	Low	CC	L. Room	N/A	N/A	N/A	RGB-D	Egocentric
Xie and Dames [59]	Medium	N/A	Field	354x354x420	0.50	DD	LIDAR	Egocentric
Yang et al. [60]	Low	P, R	Hallway	700x400x500	3.30	Quad	Motion Capture	Overhead
Yao et al. [61]	Low	P, C	Hallway	354x354x420	0.65	DD	LIDAR, RGB-D	Egocentric

Sensors. Most works use limited sensing capabilities, leading to a more challenging evaluation. We note that some of the potential negative effects of limited sensing (false detections, occlusion) are more problematic in scenarios with higher numbers of pedestrians, and might not be stress-tested in many of the evaluations we covered.

Viewpoint. We differentiate between *overhead*, in which sensors have a bird's eye view, mitigating issues like occlusions, and *egocentric*, in which sensors are on the robot.

IV. Investigating the Complexity of Social Robot Navigation Scenarios

We propose a definition of a *Social robot navigation scenario* based on parameters that concisely capture the contextual factors in Sec. III. We then describe an experiment design that investigates the performance of a variety of navigation algorithms under different scenarios.

A. Social Navigation Scenario

Consider a robot navigating next to $n \geq 1$ human agents in a workspace $\mathcal{W} \subseteq \mathbb{R}^2$ with a set of static obstacles $\mathcal{W}_{obs} \subseteq \mathcal{W}$. The robot starts from an initial configuration s_R and moves towards a goal g_R by following a policy π_R whereas humans are navigating from their initial configurations s_i towards their goals, g_i by following a policy π_i, $i \in \mathcal{N}$; agents' goals are unknown to one another. The robot occupies an area $A_R \in \mathcal{W}$, and each human occupies an area $A_i \in \mathcal{W}$. The objective of the robot is to reach its destination while avoiding collisions and abiding by social norms. We define a social navigation scenario as a tuple:

$$\mathcal{S} = (n, A_R, A_{1:n}, s_R, g_R, s_{1:n}, g_{1:n}, \pi_{1:n}, \mathcal{W}_{obs}) . \quad (1)$$

We denote by π_i the true policy for agent i, capturing the way they make decisions based on their objectives as well as behavioral and contextual aspects of their navigation profile.

B. Experiment Design

We design scenarios of varying Complexity by manipulating each of the factors in Sec. III in isolation.

Scenario configurations. We first define a base scenario $\mathcal{S}_b$ with $n = 15$, $A_R, A_{i:n} = \pi(0.3)^2$, $s_R = (5, 1)$, $g_R = (5, 9)$, $\pi_{i:n}$ = *ORCA, SFM*, $v_{pref} = 1.0$ m/s, $\mathcal{W}_{obs} = \{[0, w] \times [0, l]\}^c = \{[0, 10] \times [0, 10]\}^c$; $s_{1:n}, g_{1:n}$ are sampled from passing and crossing. We then modify a single variable in each experiment, from low to high *intensity*:

- *Density*: {0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35}. We choose this range following prior work [1, 9].

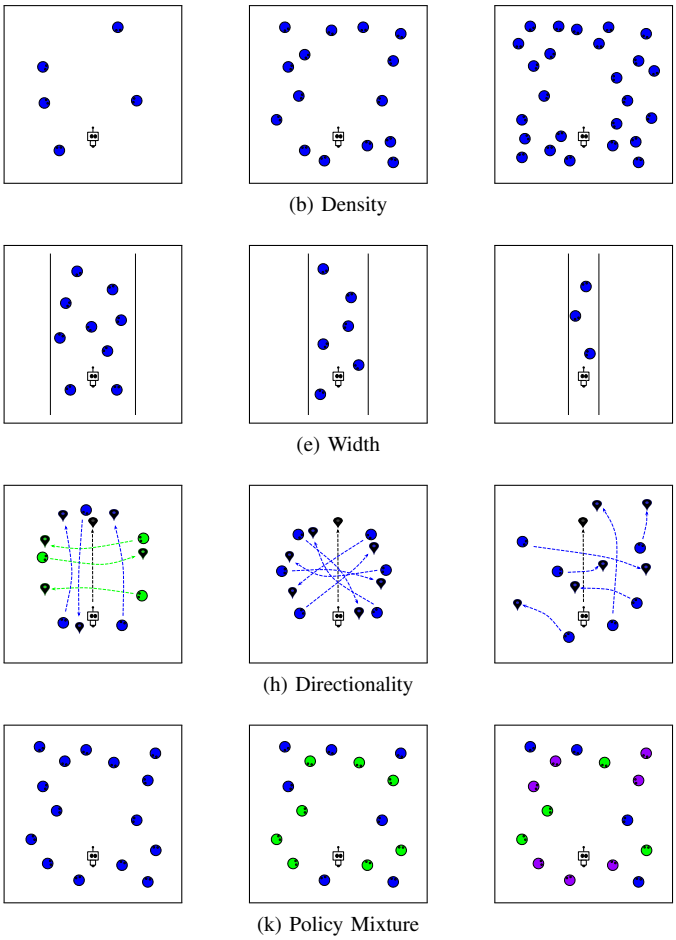


Fig. 2: Left to right: Scenarios of increasing Complexity for each of the Complexity factors considered (b-k). In the leftmost Directionality figure green agents are crossing and blue agents are passing, while different colors represent different policies in Policy Mixture.

- *Directionality*: {Passing only, crossing only, passing and crossing, circle crossing, random start/goal}. We found that prior evaluations frequently involve passing scenarios with one or two humans, although many papers set up simultaneous passing and crossing scenarios. We also find that while circle crossings are frequently used in simulation to force Complex interactions between several agents [1, 64–66], they do not appear as often in real robot evaluations.
- *Policy Mixture*: {SFM only, ORCA only, Mix 1, Mix 2, Mix 3}, where Mix 1 contains 8 ORCA and 7 SFM agents, Mix 2 contains 5 ORCA, 5 SFM, 2 CV, 3 Static agents, and Mix 3 contains 4 ORCA, 4 SFM, 4 CV, 3 Static agents, respectively. We add increasing numbers of inattentive agents to model Complex real-world scenarios in which the robot must navigate among both cooperative and uncooperative agents simultaneously [41].
- *Environment Width*: {4.5, 4.0, 3.5, 3.0, 2.5, 2.0, 1.5}. which reflects the reported Width of most hallways (e.g., [17, 39, 60]) in Table. I.

Fig. 2 illustrates the configurations considered in our study.

Algorithms. We propose to evaluate the change in Complexity of a given scenario using the performance of several navigation algorithms. Specifically, we employ:

- *Relational Graph Learning (RGL) [64]:* a reinforcement learning based approach that models pedestrian interactions using a graph neural network.
- *Social GAN Model Predictive Control (MPC-SGAN):* An approach integrating a recent crowd motion prediction model [67] into an MPC framework, using the approach and implementation of Poddar et al. [16].
- *Constant Velocity Model Predictive Control (MPC-CV):* Identical to *MPC-SGAN*, but instead using CV motion predictions for each agent.
- *Social GAN Model Predictive Path Integral (MPPI-SGAN):* SGAN integrated into an MPPI controller.
- *Reactive Planner (RP)*: A myopic planner that attempts to avoid collisions while navigating towards the goal.
- *Optimal Reciprocal Collision Avoidance (ORCA) [63]*: A multiagent collision avoidance method that guarantees collision free movement among ORCA agents.
- *Social Forces Model (SFM) [23]*: A physics-inspired model of crowd motion.
- *Constant Velocity (CV)*: An unreactive agent that moves with constant velocity toward the goal.

Metrics. Based on literature [1, 2, 10], we use the following metrics for performance, collected from Successful trials:

- *Success:* The average number of trials in which the ego-agent reaches its goal without collision.
- *Time to goal:* The average time to goal across trials.
- *Distance to agent:* The Minimum Distance to the nearest other agent during a trial, averaged across trials.
- *Path Irregularity [31]:* The amount of unnecessary turning per unit path length, measured in $\frac{rad}{m}$, calculated as $\sum_{Path} \frac{Rotation - Min.\ rotation\ needed}{Path\ length}$.

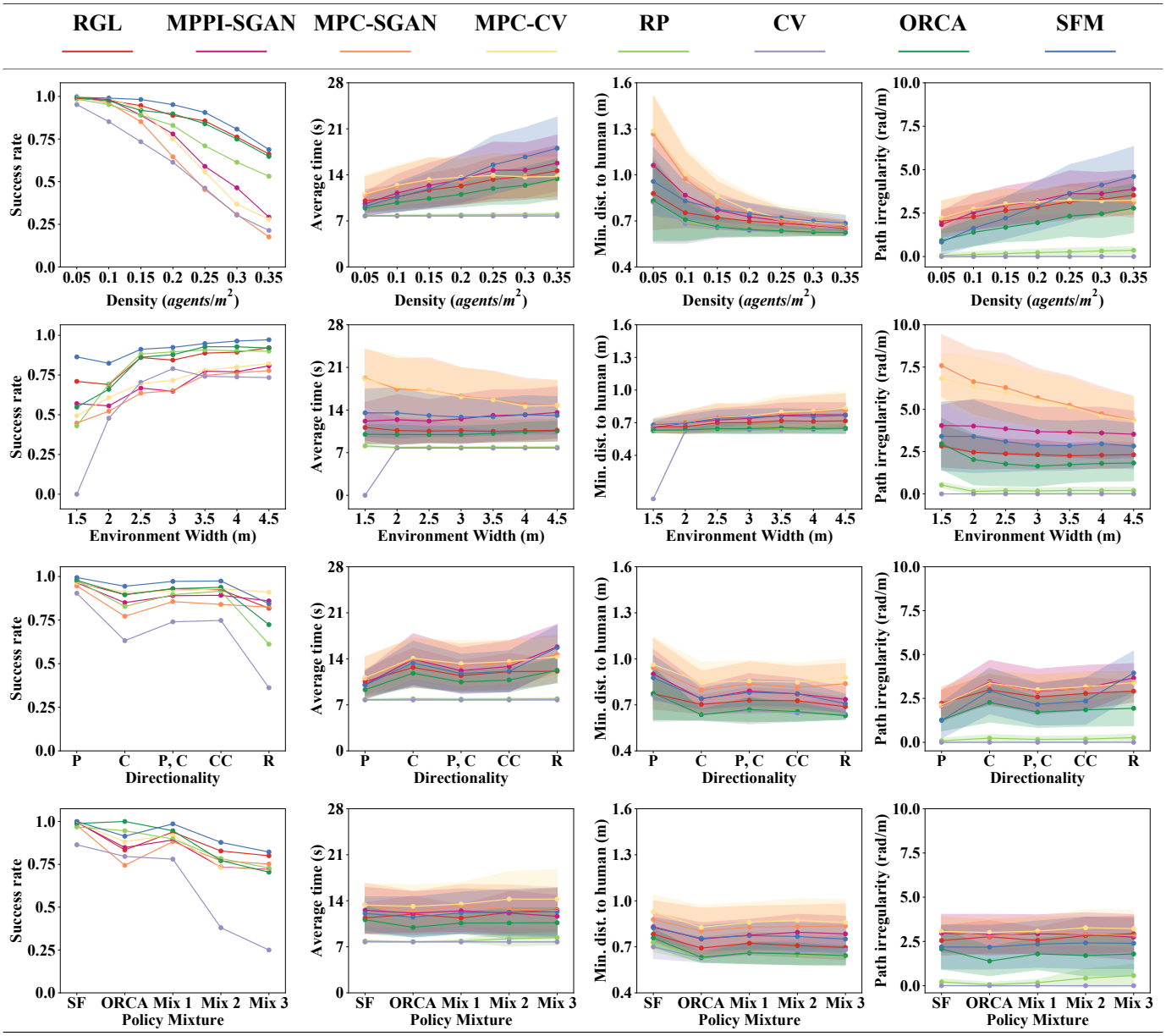
Hypotheses. We expect that scenarios of higher Complexity will pose greater navigation challenges, and this will be reflected in significant performance drops across all algorithms. Additionally, as Density is often used as a proxy for Complexity as a whole, we anticipate that it will have the strongest correlation with observed performance drops. Based on these expectations, We formulate the following hypotheses:

- *H1*. Increasing the intensity of each of the four Complexity factors (Density, Directionality, Environment Geometry, Policy Mixture) independently decreases performance with respect to (w.r.t.) collected metrics.
- *H2*. The negative correlation between Complexity and performance in *H1* will be strongest for Density.

C. Implementation Details

We generate 500 scenarios for each condition within each experiment and report metrics as averages across all 500 trials. We fix the random seed to ensure each method experiences the same scenarios. To simulate continuous crowd motion, agents each have a precomputed sequence of goals sampled according to the Directionality of the scenario, and begin moving to their next goal upon reaching their current one. For MPC-SGAN and MPPI-SGAN, we used a checkpoint pretrained on the *Zara* portion of the UCY dataset [68]. RGL was retrained in 5, 10, and 15 agent scenarios with ORCA, SFM, and

Fig. 3: Performance of methods across our experiments. Rows indicate experiments and columns correspond to different evaluation metrics. Each point represents the mean over 500 experiments; shaded regions indicate standard deviation. Mix 1 has 7 SFM and 8 ORCA agents. Mix 2 has 5 SFM, 5 ORCA, 2 CV, and 3 static agents. Mix 3 has 4 SFM, 4 ORCA, 4 CV, 3 static agents.



ORCA/SFM agent policies, however the training only resulted in a collision-avoiding policy in the 5 ORCA agent setting. For the ORCA and SFM methods, we used the CrowdNav default parameter settings. The reactive planner uses the same action space of Chen et al. [64], and selects the collision free action that minimizes its distance to the goal in the following timestep. All methods handle static obstacle collisions by setting the component of the action in the direction of the colliding obstacle to 0. For more specific details of the training procedure for RGL, and parameter settings and tuning for the MPC and MPPI methods, as well as our enhanced version of CrowdNav, see our GitHub site at https://github.com/fluentrobotics/ComplexityNav. Videos of our experiments can be found at https://youtu.be/-ir12VoSCkY.

D. Results

Fig. 3 summarizes our experimental results, organized by experiment types (rows) and metrics (columns). Based on this data, we investigate the validity of our hypotheses:

H1. We find Density has a strong correlation with Success Rate and Minimum Distance to agent ($\rho = -0.878, -0.760$, $p < 0.001$ using Spearman's r test), and a moderate correlation with Average Time ($\rho = 0.488$, $p < 0.001$). We also find a strong correlation between *increasing* Width and Success Rate ($\rho = 0.641$, $p < 0.001$), and a moderate correlation between Width and Minimum Distance ($\rho = 0.420$, $p < 0.01$). We do not find any statistically significant correlation with average Time or Path Irregularity. Regarding Policy Mixture and Directionality, we find strong and moderate correlations respectively between intensity and Success Rate ($\rho = -0.778, -0.524$, $p < 0.002$). We see no statistically significant correlation for

Minimum Distance and Time in either case, and no statistically significant correlations for Path Irregularity. Thus, we find that while each factor affects at least one metric, Density and Environment Geometry appear to have the strongest correlation, giving partial support for *H1*.

H2. We observe that the correlations between Density and other collected metrics are all stronger than those of the other factors. Thus we find support for *H2*.

E. Analysis

While Density correlates strongest with Complexity, we see that the Environment Geometry, Directionality, and agent policies all have at least some correlation with performance, and thus should be considered when assessing the Complexity of a scenario. While the support for **H2** vindicates prior use of Density as a proxy for Complexity, the partial support for **H1** suggests that a more accurate picture can be obtained by analyzing the rest of the contextual factors.

Despite being trained on only 5 ORCA agent circle crossings, RGL achieves higher Success than the MPC and MPPI methods in nearly all experiments, particularly in high Density scenarios. Its Success Rate nearly mirrors ORCA (its training Policy Mixture), however with consistently higher Minimum Distances to humans. This indicates it learned similar collision avoidance behavior to ORCA, while also learning to better respect agents' personal space.

MPC-CV, MPC-SGAN, and MPPI-SGAN performance scales poorly with moderate to high Density, although we do see MPPI-SGAN scales better than MPC-SGAN, showing that a stronger controller implementation does indeed lead to better performance (albeit with slightly lower Minimum Distances). The steep decrease for all three methods, however, indicates that having an inaccurate prediction model becomes increasingly problematic as the number of agent trajectories predicted increases, even with a more robust controller.

In nearly all experiments, RP and ORCA policies had low average Time, while still maintaining comparable or better Success rates to other methods. Even the CV agent was moderately Successful in lower Complexity scenarios, which matches prior experimental outcomes [1]. SFM's high Success Rates and Minimum Distances indicate that with proper tuning it could be viable as a local controller for a socially navigating robot, although more must be done to evaluate whether its effectiveness transfers out of simulation.

We do however see that the MPC methods maintain the highest Minimum Distance, while ORCA, RP, and CV have the lowest Minimum Distances. Additionally, the MPC methods, with the exception of the Environment Width experiments, maintain comparable Path Irregularity to ORCA, SFM, and RGL, showing they can maintain distance with comparably smooth paths. Thus we see that while the reactive and CV methods (with the exception of SFM) achieve efficient, collision-free navigation among cooperative agents, they are worse at respecting the personal space of others compared to predictive methods.

We obtained evidence that increased heterogeneity of scenarios is not by itself an indicator of increased Complexity. Qualitatively, we observed that passing scenarios are generally easier to navigate than crossing, and SFM agents are much more subservient than ORCA. An implicit hypothesis in our experimental ordering was that a mixture of two Directionalities or Policy Mixtures would be more Complex to navigate than either individually. We instead see improvement in the ORCA/SFM and Passing/Crossing scenarios compared to ORCA and Crossing only, which suggests the combinations are actually less Complex.

V. Discussion

Conducting high-Complexity evaluations. Evaluations in the literature are often conducted under low-Density settings (Table I). Our experiments demonstrate that most algorithms handle those well but experience significant performance reductions as Density increases. Similarly, while many works contain passing-only experiments, in our study, Passing has the highest Success rates and Minimum Distances to agents, indicating it is the easiest to navigate. Additionally, while many works consider Medium-to-Large rooms, these appear to be the simplest to navigate (Fig. 3). Finally, nearly all evaluations in recent literature focus on scenarios with exclusively cooperative agents, which our experiments show are easiest. These practices hinder the extraction of helpful insights; for example, had we not manipulated the Complexity factors towards the upper extremes, the severe performance drops at high Density experienced by MPC-CV, MPC-SGAN, and MPPI-SGAN would not have been identified. These observations suggest that the social navigation community should shift towards studying scenarios of higher Complexity to move beyond the frontier. Specifically, more evaluations in more geometrically-constrained, highly-mixed and high-traffic environments, similar to many context-rich public domains should be conducted.

Handling test-time distribution. The high Success rates of CV and RP suggest that in lower-Complexity regimes, it might be sufficient to use non-reactive classical planning techniques. As the Complexity increases, all algorithms experience a steep performance decline, the magnitude of which serves to indicate how far out of distribution an algorithm operates. Our evaluation (Fig. 3) captures the sensitivity of data-driven approaches to their training distribution. For instance, we see that RGL, and MPC/MPPI-SGAN experience a substantial performance decline as the test-time Density moves away from the training/tuning Density ($0.05, 0.10\ agents/m^2$). In practice, data-driven approaches will often face distribution shift at deployment, leading to unexpected or unsafe behavior, as demonstrated by the performance drop experienced by RGL and the MPC approaches in scenarios of different Complexity than their training data. Thus, to safely and effectively deploy social navigation algorithms, it will be important to rigorously characterize the scenarios in which they can be expected to perform effectively, and which scenarios are out of distribution and may cause performance degradation.

Reproducing Complexity. In our literature review, we found that the amount of details provided varied substantially among papers, leaving us to approximate values based on

figures or supplementary material, when available. We believe that sharing more precise details related to the Complexity factors we identified will be an important step towards better reproducibility in social navigation.

Simulating complex settings. While realistically simulating pedestrian-robot interactions is challenging [1, 29], we view it as an essential tool for testing high-Complexity settings that are difficult to safely replicate in the real world. This requires revisiting conventional assumptions, such as that humans are non-reactive to the robot [13, 36, 42, 43, 58, 64], which is unrealistic since the robots used (see Table I) generally have a large enough footprint to be observed. Thus, simulation should focus on *visible* robot settings, leveraging metrics and considerations of users' perceptions [26–28] to compare algorithms' social performance.

Limitations. Our analysis is a step towards understanding the Complexity of social navigation scenarios; an important next step would be to validate our observations with real-world experiments. While our evaluation involves standard metrics [1, 2], it lacks a user-centered perspective. Future work will incorporate metrics mapping robot behavior to users' impressions [26, 27, 69]. We tested algorithms for which we found open-source implementations but additional algorithms could be included. We strived for a fair comparison by tuning all algorithms w.r.t. the same criteria. While tuning can always be improved, the key takeaway remains: there is a steep drop in performance as Complexity increases. Additionally, we acknowledge that our simulator makes the unrealistic assumptions of perfect sensing and constant-sized circular agents, and these could be improved in future implementations. Furthermore, our results are affected by our choice of ORCA and SFM to simulate cooperative agents, where other algorithms may have led to different results. While these assumptions reduce the realism of our experiments, they allow us to minimize the effect of extraneous factors on the relationship between the Complexity factors and performance. Future work will investigate the robot-related factors of kinematics, sensors, and viewpoints, and can explore the effects of additional real-world considerations.

REFERENCES

[1] C. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, A. Steinfeld, P. Trautman, and J. Oh, "Core challenges of social robot navigation: A survey," *ACM Transactions on Human-Robot Interaction*, vol. 12, pp. 1 – 39, 2021.

[2] A. Francis et al., "Principles and guidelines for evaluating social robot navigation algorithms," *arXiv:2306.16740* [cs.RO], 2023.

[3] P. T. Singamaneni, P. Bachiller-Burgos, L. J. Manso, A. Garrell, A. Sanfeliu, A. Spalanzani, and R. Alami, "A survey on socially aware robot navigation: Taxonomy and future challenges," *The International Journal of Robotics Research*, vol. 0, no. 0, p. 02783649241230562, 2024.

[4] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*. Cambridge University Press, 2006.

[5] V. Vapnik, *Statistical learning theory*. Wiley New York, 1998.

[6] M. Farber, "Topological complexity of motion planning," *Discrete & Computational Geometry*, vol. 29, no. 2, pp. 211–221, 2003.

[7] D. E. Koditschek, "What is Robotics? Why Do We Need It and How Can We Get It?" *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 1–33, 2021.

[8] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixe, "CVPR19 Tracking and Detection Challenge: How crowded can it get?" 2019.

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

[10] Y. Gao and C.-M. Huang, "Evaluation of socially-aware robot navigation," *Frontiers in Robotics and AI*, vol. 8, 01 2022.

[11] L. Tai, J. Zhang, M. Liu, and W. Burgard, "Socially compliant navigation through raw depth inputs with generative adversarial imitation learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1111–1117.

[12] S. Liu, P. Chang, W. Liang, N. Chakraborty, and K. Driggs-Campbell, "Decentralized Structural-RNN for robot crowd navigation with deep reinforcement learning," in *in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 3517–3524.

[13] S. Liu, P. Chang, Z. Huang, N. Chakraborty, K. Hong, W. Liang, D. L. McPherson, J. Geng, and K. Driggs-Campbell, "Intention aware robot crowd navigation with attention-based interaction graph," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 015–12 021.

[14] C. I. Mavrogiannis, W. B. Thomason, and R. A. Knepper, "Social momentum: A framework for legible navigation in dynamic multi-agent environments," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 361–369.

[15] C. Mavrogiannis, K. Balasubramanian, S. Poddar, A. Gandra, and S. S. Srinivasa, "Winding through: Crowd navigation via topological invariance," *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 121–128, 2023.

[16] S. Poddar, C. Mavrogiannis, and S. S. Srinivasa, "From crowd motion prediction to robot navigation in crowds," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 6765–6772.

[17] E. Pacchierotti, H. I. Christensen, and P. Jensfelt, "Evaluation of passing distance for social robots," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2006, pp. 315–320.

[18] S. Lo, K. Yamane, and K. Sugiyama, "Perception of pedestrian avoidance strategies of a self-balancing mobile robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 1243–1250.

[19] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: a second-generation museum tour-guide robot," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 3, 1999, pp. 1999–2005.

[20] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 1998, pp. 11–18.

[21] A. Biswas, A. Wang, G. Silvera, A. Steinfeld, and H. Admoni, "SocNavBench: A grounded simulation testing framework for evaluating social navigation," *Transactions on Human-Robot Interaction*, vol. 11, no. 3, 2022.

[22] J. van den Berg, S. J. Guy, M. Lin, and D. Manocha, "Reciprocal n-Body collision avoidance," in *Robotics Research*. Springer Berlin Heidelberg, 2011, pp. 3–19.

[23] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.

[24] J. Holtz and J. Biswas, "Socialgym: A framework for benchmarking social robot navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 246–11 252.

[25] N. Tsoi, A. Xiang, P. Yu, S. S. Sohn, G. Schwartz, S. Ramesh, M. Hussein, A. W. Gupta, M. Kapadia, and M. Vázquez, "SEAN 2.0: Formalizing and generating social situations for robot navigation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 047–11 054, 2022.

[26] N. Walker, C. Mavrogiannis, S. S. Srinivasa, and M. Cakmak, "Influencing behavioral attributions to robot motion during task execution," in *Proceedings of the Conference on Robot Learning (CoRL)*, 2021.

[27] C. Mavrogiannis, P. Alves-Oliveira, W. Thomason, and R. A. Knepper, "Social momentum: Design and evaluation of a framework for socially competent robot navigation," *Transactions on Human-Robot Interaction*, vol. 11, no. 2, 2022.

[28] P. T. Singamaneni, A. Favier, and R. Alami, "Towards benchmarking human-aware social robot navigation: A new perspective and metrics," in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 914–921.

[29] T. Fraichard and V. Levesy, "From crowd simulation to robot navigation in crowds," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 729–735, 2020.

[30] M. Sun, F. Baldini, P. Trautman, and T. Murphey, "Move beyond trajectories: Distribution space coupling for crowd navigation," in *Proceedings of Robotics: Science and Systems (RSS)*, 2021.

[31] J. Guzzi, A. Giusti, L. M. Gambardella, and G. Di Caro, "Human-friendly robot navigation in dynamic environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 05 2013, pp. 423–430.

[32] I. Dynnikov and B. Wiest, "On the complexity of braids," *Journal of the European Mathematical Society*, vol. 9, no. 4, pp. 801–840, 2007.

[33] G. Angelopoulos, A. Rossi, C. D. Napoli, and S. Rossi, "You are in my way: Non-verbal social cues for legible robot navigation behaviors," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 657–662.

[34] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning motion patterns of people for compliant robot motion," *The International Journal of Robotics Research*, vol. 24, no. 1, pp. 31–48, 2005.

[35] P.-H. Ciou, Y.-T. Hsiao, Z.-Z. Wu, S.-H. Tseng, and L.-C. Fu, "Composite reinforcement learning for social robot navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2553–2558.

[36] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *Procedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 285–292.

[37] K. Katyal, Y. Gao, J. Markowitz, S. Pohland, C. Rivera, I.-J. Wang, and C.-M. Huang, "Learning a group-aware policy for robot navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 328–11 335.

[38] B. Kim and J. Pineau, "Socially adaptive path planning in human environments using inverse reinforcement learning," *International Journal of Social Robotics*, vol. 8, no. 1, pp. 51–66, 2016.

[39] R. Kirby, "Social robot navigation," Ph.D. dissertation, Carnegie Mellon University, 2010.

[40] L. Liu, D. Dugas, G. Cesari, R. Siegwart, and R. Dubé, "Robot navigation in crowded environments using deep reinforcement learning," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5671–5677.

[41] S. Matsuzaki, S. Aonuma, and Y. Hasegawa, "Dynamic window approach with human imitating collision avoidance," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 8180–8186.

[42] S. Matsuzaki and Y. Hasegawa, "Learning crowd-aware robot navigation from challenging environments via distributed deep reinforcement learning," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2022, pp. 4730–4736.

[43] Y.-J. Mun, M. Itkina, S. Liu, and K. Driggs-Campbell, "Occlusion-aware crowd navigation using people as sensors," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 031–12 037.

[44] K. A. Mustafa, O. de Groot, X. Wang, J. Kober, and J. Alonso-Mora, "Probabilistic risk assessment for chance-constrained collision avoidance in uncertain dynamic environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 3628–3634.

[45] D. M. Nguyen, M. Nazeri, A. Payandeh, A. Datar, and X. Xiao, "Toward human-like social robot navigation: A large-scale, multi-modal, social human navigation dataset," 2023.

[46] J. Oh, J. Heo, J. Lee, G. Lee, M. Kang, J. Park, and S. Oh, "SCAN: Socially-aware navigation using monte carlo tree search," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 7576–7582.

[47] D. Paez-Granados, Y. He, D. Gonon, D. Jia, B. Leibe, K. Suzuki, and A. Billard, "Pedestrian-robot interactions on autonomous crowd navigation: Reactive control methods and evaluation metrics," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 149–156.

[48] R. Peddi, C. D. Franco, S. Gao, and N. Bezzo, "A data-driven framework for proactive intention-aware motion planning of a robot in a human environment," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5738–5744.

[49] L. Qin, Z. Huang, C. Zhang, H. Guo, M. Ang, and D. Rus, "Deep imitation learning for autonomous navigation in dynamic pedestrian environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 4108–4115.

[50] Q. Qiu, S. Yao, J. Wang, J. Ma, G. Chen, and J. Ji, "Learning to socially navigate in pedestrian-rich environments with interaction capacity," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 279–285.

[51] A. J. Sathyamoorthy, J. Liang, U. Patel, T. Guan, R. Chandra, and D. Manocha, "DenseCAvoid: Real-time navigation in dense crowds using anticipatory behaviors," *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11 345–11 352, 2020.

[52] M. Shiomi, F. Zanlungo, K. Hayashi, and T. Kanda, "Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model," *International Journal of Social Robotics*, vol. 6, pp. 443–455, 08 2014.

[53] S. Silva, N. Verdezoto, D. Paillacho, S. Millan-Norman, and J. D. Hernández, "Online social robot navigation in indoor, large and crowded environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 9749–9756.

[54] P. T. Singamaneni, A. Favier, and R. Alami, "Watch out! there may be a human. addressing invisible humans in social navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 344–11 351.

[55] X.-T. Truong and T. D. Ngo, "Toward socially aware robot navigation in dynamic and crowded environments: A proactive social motion model," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 4, pp. 1743–1760, 2017.

[56] C.-E. Tsai and J. Oh, "A generative approach for socially compliant navigation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 2160–2166.

[57] W. Wang, R. Wang, L. Mao, and B.-C. Min, "NaviSTAR: Socially aware robot navigation with hybrid spatio-temporal graph transformer and preference learning," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 11 348–11 355.

[58] R. Wang, W. Wang, and B.-C. Min, "Feedback-efficient active preference learning for socially aware robot navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 11 336–11 343.

[59] Z. Xie and P. Dames, "DRL-VO: Learning to navigate through crowded dynamic scenes using velocity obstacles," *IEEE Transactions on Robotics*, vol. 39, no. 4, p. 2700–2719, 2023.

[60] B. Yang, J. Jiao, L. Wang, and M. Liu, "An online interactive approach for crowd navigation of quadrupedal robots," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 13 556–13 562.

[61] S. Yao, G. Chen, Q. Qiu, J. Ma, X. Chen, and J. Ji, "Crowd-aware robot navigation for pedestrians with multiple collision avoidance strategies via map-based deep reinforcement learning," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 8144–8150.

[62] A. Wang, C. Mavrogiannis, and A. Steinfeld, "Group-based motion prediction for navigation in crowded environments," in *Proceedings of the Conference on Robot Learning (CoRL)*, 2021.

[63] J. van den Berg, S. J. Guy, M. C. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *Proceedings of the International Symposium on Robotics Research (ISRR)*, 2009, pp. 3–19.

[64] C. Chen, S. Hu, P. Nikdel, G. Mori, and M. Savva, "Relational graph learning for crowd navigation," in *Proceesings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 007–10 013.

[65] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6015–6022.

[66] M. M. Sun, F. Baldini, K. Hughes, P. Trautman, and T. Murphey, "Mixed strategy nash equilibrium for crowd navigation," 2024.

[67] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 2255–2264.

[68] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by Example," *Computer Graphics Forum*, 2007.

[69] M. Dadvar, K. Majd, E. Oikonomou, G. Fainekos, and S. Srivastava, "Joint communication and motion planning for cobots," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2022, pp. 4771–4777.